CONTENTS OF VOLUME 29

ARTICLES

MISCELLANEOUS

Index of Abstracts for 1958

PÓLYA TYPE DISTRIBUTIONS IV. SOME PRINCIPLES OF SELECTING A SINGLE PROCEDURE FROM A COMPLETE CLASS[1]

By Samuel Karlin

Stanford University

0. Introduction. In previous publications [1], [2], and [3], various aspects of decision theory in which the underlying distributions are Pólya type have been studied. For example, complete classes of decision procedures were determined, all Bayes procedures were characterized, and the problem of admissibility was investigated as related to various kinds of loss functions.

Usually the minimal complete class of decision procedures, to which the statistician would obviously restrict himself in practical application, is still quite large. Consequently, without any additional knowledge or further conditions, it is a hopeless task to justify preferring any given admissible procedure to another. It is therefore of importance to introduce new criteria which will single out a procedure for use. It is the object of this paper to discuss some further principles which select a single statistical procedure from the class of all "monotone" procedures.

In the $n = 2$ action problem (essentially the testing problem) some of the classical principles used to determine a single admissible procedure for use are related to the concepts of unbiasedness, maximum likelihood, invariance, minimax, etc. These principles have received much attention and their justification and relevance are well understood for the parametric testing problem. For a detailed analysis of these classical concepts in the case of two action problems when the underlying distributions are Pólya type, the reader is referred to [1]. Our present discussion deals with the extension and analysis of some of these principles to the n-action problem. In the sense that the estimation problem may be obtained as a limit of finite action problems, the ideas here shed further light on the estimation problem.

The language and notation we use is that of the introduction of the previous paper [3]. However, a knowledge of the results of [3] is not necessary for an understanding of the present discussion although a reading of the introduction would more than provide sufficient familiarity with the terminology to be used here as well as a general background for Pólya type distributions. Henceforth, we assume that the notation of this manuscript is that of [3]. Nevertheless, for clarity of exposition, we review briefly some of the main quantities to be used.

Let the distribution of the observed real random variable X (usually a sufficient

Received March 25, 1957.

[1] Research sponsored by the Office of the Naval Research Nonr 225 (21) (NR 042-993). Research on the paper "Pólya-Type Distributions III: Admissibility of Multi-Action Problems," in the December, 1957 issue of these *Annals*, was done under the same contract.

statistic), depending on the unknown parameter ω ($\omega \varepsilon \Omega$, an interval of the real line), have the form

$$P(x, \omega) = \int_{-\infty}^{x} p(\xi, \omega)\, d\mu(\xi), \tag{1}$$

where the density $p(\xi, \omega)$ possesses a monotone likelihood ratio (Pólya type 2) and μ is a countably additive measure defined at least for the Borel field of sets containing the open subsets of the real line. Occasionally, we shall assume the stronger condition that the density is Pólya type 3.

The main transformation property of Pólya type 2 densities used in our analysis is as follows: If $g(x)$ changes sign at most once (say from negative to positive values), then

$$h(\omega) = \int g(x)p(x, \omega)\, d\mu(x)$$

changes sign at most once. Moreover, if $h(\omega)$ does indeed change signs, then it must change in the same direction as g, i.e., from negative to positive. For a thorough discussion of these properties the reader is referred to [2].

There are n possible actions, and $L_i(\omega)$ ($i = 1, \cdots, n$) represents the measure of the loss when taking action i and ω is the state of nature. We require that the set

$$S_i = \{\omega \mid L_i(\omega) < L_j(\omega), j \neq i\} = (\omega_{i-1}^0, \omega_i^0)$$

where the ω_i^0 satisfy

$$-\infty = \omega_0^0 < \omega_1^0 < \omega_2^0 < \cdots < \omega_n^0 = \infty.$$

The set S_i represents the set of ω values where action i is favored if the state of nature were known. Also, we assume that $L_i(\omega) - L_{i+1}(\omega)$ has exactly one sign change which must occur at ω_i^0.

We shall assume throughout what follows that the loss functions $L_i(\omega)$ and the density $p(x, \omega)$ satisfy sufficient smoothness conditions to guarantee the existence of all integrals involving these quantities and to justify all differentiation operations. In most particular examples these smoothness requirements can be readily verified.

A statistical procedure is an n-tuple

$$\varphi(x) = (\varphi_1(x), \cdots, \varphi_n(x)),$$

where $\varphi_i(x)$ is interpreted as the probability of taking action i when observing x. A "monotone" procedure is characterized by a tuple

$$(x_1, x_2, \cdots, x_{n-1}; \lambda_1, \cdots, \lambda_{n-1})$$

where $x_1 \leqq x_2 \leqq \cdots \leqq x_{n-1}$, $0 \leqq \lambda_i \leqq 1$. Explicitly, when the x_i are distinct, then

$$\varphi_i(x) = \begin{cases} 1 & \text{if } x_{i-1} < x < x_i, \\ 0 & \text{if } x < x_{i-1}, x > x_i, \\ \lambda_i & \text{if } x = x_i, \\ 1 - \lambda_{i-1} & \text{if } x = x_{i-1}; \end{cases} \qquad i = 1, \cdots, n,$$

and by definition $x_0 = -\infty$, $\lambda_0 = 0$, $x_n = +\infty$, $\lambda_n = 1$ In the case where some of the x_i coincide then appropriate changes in the form of the definition of $\varphi_i(x)$ at the values x_i must be made. If the measure μ of (1) has no atoms (jumps), then a monotone procedure is fully specified (up to equivalence almost everywhere with respect to μ) by the critical values $(x_1, x_2, \cdots, x_{n-1})$. For the sake of simplicity of exposition, we restrict ourselves henceforth to the case of a continuous distribution. However, we remark in passing that all of the results of this paper may be extended, subject to suitable modifications, to the general case where we allow the measure μ to possess atoms. The risk corresponding to any given strategy $\varphi = (\varphi_1, \varphi_2, \cdots, \varphi_n)$ is given by the expression

$$\rho(\omega, \varphi) = \int p(x, \omega) \left\{ \sum_{i=1}^{n} L_i(\omega)\varphi_i(x) \right\} d\mu(x). \tag{2}$$

The collection of all monotone procedures constitutes a complete class [4]. When the loss functions satisfy additional assumptions, then all non-degenerate monotone procedures are also admissible [3]

The set of all monotone strategies $\mathfrak{M}$ form an $n - 1$ dimensional family in the sense that they depend on the $n - 1$ critical values which determine the procedures. Our problem, in choosing a specific strategy from $\mathfrak{M}$, is in essence finding $n - 1$ conditions which will cut the class $\mathfrak{M}$ down to a unique member. Alternatively, we could impose some global restrictions which also single out a monotone procedure. For instance, if an a priori distribution of nature $F(\omega)$ is known to be meaningful, then the Bayes procedure with respect to F determines a specific monotone procedure. [See [3], [5].] The assumption of the existence of F is often hard to justify and appears contrived.

Another global condition frequently followed is to choose a monotone minimax procedure. However, minimax procedures are often very unreasonable on the basis of statistical intuition and there exists feeling that minimax philosophy is in general too conservative and unrealistic. Of course, modifications of the minimax principle lead to the so-called regret principles. Various complications appear also for the case of the criteria of minimax regret [6].

A third method for choosing a monotone procedure is inherent in the construction of complete classes as introduced in [4]. Suppose that for a given problem there has been in use a common or accepted mode of action which is not a monotone procedure. Then, there exists at least one monotone procedure which improves everywhere on it for the decision problem of more than two actions. If the original procedure is described by an n-tuple of functions $\varphi = (\varphi_1, \cdots, \varphi_n)$, then any monotone procedure $\varphi^0 = (\varphi_1^0, \cdots, \varphi_n^0)$ (and there is at least one) which satisfies

$$\int_{-\infty}^{\infty} p(x \mid \omega) \left[\sum_{j=1}^{i} \varphi_j^0(x) - \sum_{j=1}^{i} \varphi_j(x) \right] d\mu(x) \begin{cases} \geqq 0 & \text{for} \quad \omega \leqq \omega_i^0 \\ \leqq 0 & \text{for} \quad \omega \geqq \omega_i^0 \end{cases}$$

improves on φ.

This method is constructive. That is, for any non-monotone procedure in use we can explicitly exhibit a monotone procedure which yields a smaller risk uni-

formly for any choice of the state of nature ω. The apparent disadvantage to this idea is that it involves only an improvement relative to a given non-monotone procedure and sheds no light on the intrinsic question of selecting a specific monotone procedure from the class $\mathfrak{M}$.

In this study we will analyze three principles of selecting a single monotone procedure from $\mathfrak{M}$. The first represents an extension of the maximum likelihood estimate to the circumstance of the n-action problem. The monotone test obtained in this case has a lot of intuitive appeal and will be referred to as the maximum likelihood procedure.

The following section examines another approach called the principle of maximum probabilities (abbreviated M.P.). This principle, as well as the maximum likelihood procedure, does not depend on the specific values of the loss functions but rather on the preference regions $S_i = (\omega_{i-1}^0\, \omega_i^0)$. Any other loss function satisfying the properties of a monotone preference pattern and giving rise to the same preference sets S_i will possess the same class of monotone procedures obeying the principle of M.P.

The precise description of this principle is as follows: A decision procedure which is defined by an n-tuple of functions $\varphi = (\varphi_1, \cdots, \varphi_n)$ is said to have the property of maximum probabilities (φ has M.P.) if for every i

$$h_i(\omega') \geqq h_i(\omega'') \qquad \text{for any } \omega' \text{ in } S_i,\ \omega'' \not\in S_i, \tag{3}$$

where

$$h_i(\omega) = \int \varphi_i(x) p(x, \omega)\, d\mu(x). \tag{4}$$

For the case of two actions a procedure φ has the property of M.P. if and only if φ is unbiased in the classical sense. Therefore, this principle may be considered to be a generalization to the case of n actions of the concept of unbiasedness. The quantity $h_i(\omega)$ may be interpreted as the unconditional probability for the procedure φ of taking action i when the state of nature is ω. The condition (3) states that $h_i(\omega)$ is larger when ω is in S_i than when ω is outside S_i. This last property is the reason for the name, principle of maximum probabilities.

It will be shown that there always exist monotone procedures having the property of M.P. for the case of $n \leqq 5$ actions. In fact, we shall exhibit a one parameter family of such procedures. When $n > 5$, in general there ceases to exist such monotone procedures.

The final principle investigated is the principle of unbiasedness (in the sense of Lehmann [7]). A decision procedure φ is said to be risk unbiased with respect to the loss functions L_i if $E_\theta[L(\omega, \varphi(x))] \geqq E_\theta[L(\theta, \varphi(x))]$ for all ω and θ, where $E_\theta(\cdot)$ denotes the expected value given that the state of nature is θ, and

$$L(\omega, \varphi(x)) = \sum L_i(\omega)\varphi_i(x).$$

For the case of two actions, this definition reduces to the usual concept of unbiasedness. This principle of unbiasedness differs from the principle of M.P. in

that the former depends in a very crucial way on the magnitudes of the loss functions while the latter depends only on the preference regions. We shall prove that if $L_j(\omega) = L_{ij}$ for ω in S_i and the L_{ij} satisfy suitable assumptions, then there exists a unique admissible monotone procedure unbiased in the sense of Lehmann. The method of proof of the existence will in effect be constructive. In general, risk unbiased procedures need not exist.

ACKNOWLEDGEMENT. I wish to thank Mr. R. Miller for his help in the preparation of this manuscript.

1. Maximum likelihood principle. We assume throughout this section that the density $p(x, \omega)$ of (1) has a strict monotone likelihood ratio and further that $p(x, \omega)$ possesses continuous second order partial derivatives. The fact that p is of Pólya type 2 implies (see [2]) that

$$(5) \qquad \begin{vmatrix} p(x, \omega) & \dfrac{\partial}{\partial \omega} p(x, \omega) \\ \dfrac{\partial}{\partial x} p(x, \omega) & \dfrac{\partial^2}{\partial x \partial \omega} p(x, \omega) \end{vmatrix} \geqq 0$$

for all x and ω. An additional assumption is imposed to the effect that the inequality of (5) is strict for all x and ω. Finally, we assume that for each x in X the equation

$$(6) \qquad \frac{\partial}{\partial \omega} p(x, \omega) = 0$$

has a unique solution, $\omega = \omega(x)$, which is a differentiable function of x. These assumptions are not as stringent as may appear offhand. A wide class of distributions, including the exponential family ($p(x, \omega) = e^{\omega x}\beta(\omega)$), the noncentral t, the noncentral χ^2, etc., fulfills these requirements. For the exponential family, $\omega(x)$ is the solution of the equation $-\beta'(\omega)/\beta(\omega) = x$.

LEMMA 1. *$\omega(x)$ is a strictly increasing function of x.*

Proof. Differentiating Eq. (6) with respect to x leads to

$$(7) \qquad \frac{\partial^2 p(x, \omega(x))}{\partial x \partial \omega} + \frac{\partial^2 p(x, \omega(x))}{\partial \omega^2} \omega'(x) = 0.$$

By assumption,

$$\begin{vmatrix} p(x, \omega(x)) & \dfrac{\partial}{\partial \omega} p(x, \omega(x)) \\ \dfrac{\partial}{\partial x} p(x, \omega(x)) & \dfrac{\partial^2}{\partial x \partial \omega} p(x, \omega(x)) \end{vmatrix} > 0,$$

which implies $\partial^2 p(x, \omega(x))/\partial x \partial \omega > 0$ because of (6). Since $p(x, \omega)$ assumes a maximum at $\omega = \omega(x)$, $\partial^2 p(x, \omega(x))/\partial \omega^2 \leqq 0$. Thus from (7), $\omega'(x) > 0$.

As x varies over the sample space X, $\omega(x)$ varies over the whole Ω interval. Suppose not; then there exists an ω_0 such that ω_0 is not the upper endpoint of

Ω and for $\omega > \omega_0$, $\partial p(x, \omega)/\partial\omega < 0$ for all x. (Or similarly for the lower end of Ω.) But this contradicts the fact that for all $\omega \varepsilon \Omega$,

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\omega} p(x, \omega)\, d\mu(x) = \frac{\partial}{\partial\omega} \int_{-\infty}^{\infty} p(x, \omega)\, d\mu(x) = 0.$$

Since $\omega(x)$ is a $1 - 1$ strictly monotonic mapping of X onto Ω, the inverse function ω^{-1} is well-defined. Set $x_i^0 = \omega^{-1}(\omega_i^0)$, $i = 1, \cdots, n - 1$. The maximum likelihood principle dictates that the monotone procedure which should be used is the one defined by the critical numbers $(x_1^0, \cdots, x_{n-1}^0)$. For $x \varepsilon (x_{i-1}^0, x_i^0)$, take action i, $i = 1, \cdots, n$, $x_0^0 = -\infty$ and $x_n^0 = +\infty$. This principle has the feature that for any observed x the proper action i is taken whose corresponding interval $(\omega_{i-1}^0, \omega_i^0)$ includes the maximum likelihood estimate of ω. In less precise language, that action is taken which is most likely.

2. Principle of maximum probabilities (M.P.). The principle of maximum probabilities is one type of extension of the concept of unbiasedness in hypothesis testing. Consider the n-action problem defined by the points $-\infty = \omega_0^0 < \omega_1^0 < \cdots < \omega_n^0 = +\infty$ in which action i is preferred in the interval $S_i = (\omega_{i-1}^0, \omega_i^0)$. A decision procedure which is defined by an n-tuple of functions $\varphi = (\varphi_1, \cdots, \varphi_n)$ is said to have the property M.P. if for every i, $h_i(\omega') \geqq h_i(\omega'')$ for any $\omega' \varepsilon S_i$, $\omega'' \not\varepsilon S_i$, where

$$h_i(\omega) = \int_{-\infty}^{\infty} \varphi_i(x) p(x, \omega)\, d\mu(x).$$

Our object is to try to establish the existence of monotone procedures possessing the property of M.P.

It is necessary in studying this concept to assume that the density $p(x, \omega)$ is strictly Pólya type 3, and that the equation $\partial p(x, \omega)/\partial\omega = 0$ is well-defined and has a unique solution $\omega = \omega(x)$ for each value of x. For any constants $a < b$ it is tacitly assumed that differentiation with respect to ω is valid inside the integral sign of

$$\int_a^b p(x, \omega)\, d\mu(x).$$

Also, assume that μ is a continuous measure without discrete mass points whose spectrum is an interval. This last assumption is not essential but without it additional care must be taken in handling randomizations and the lack of uniqueness of various quantities caused by gaps in the spectrum.

For the purpose of exposition our analysis is divided into a series of lemmas.

A randomized strategy is now defined by $n - 1$ points $(x_1, \cdots, x_{n-1})$. Let $i(i = 1, \cdots, n - 2)$ be fixed for the moment and define $(x_i(\alpha), x_{i+1}(\alpha))$ by the equations

$$h_{i+1}(\omega_i^0) = \int_{x_i}^{x_{i+1}} p(x, \omega_i^0)\, d\mu(x) = \alpha,$$

(8)

$$h_{i+1}(\omega_{i+1}^0) = \int_{x_i}^{x_{i+1}} p(x, \omega_{i+1}^0)\, d\mu(x) = \alpha.$$

$(x_i(\alpha), x_{i+1}(\alpha))$ are uniquely defined since by Theorem 3 of [1] there is a unique monotone strategy which improves on the non-monotone strategy $\varphi(x) \equiv \alpha$. Moreover, it is clear that $h_{i+1}(\omega') \geqq h_{i+1}(\omega'')$ for any $\omega' \varepsilon S_{i+1}$ and $\omega'' \not\varepsilon S_{i+1}$ when (8) is satisfied.

LEMMA 2. *$x_i(\alpha)$ is a monotone decreasing and $x_{i+1}(\alpha)$ is a monotone increasing function of α.*

Proof. From (8),

$$(9) \qquad \int_{x_i(\alpha)}^{x_{i+1}(\alpha)} [p(x, \omega_i^0) - p(x, \omega_{i+1}^0)]\, d\mu(x) = 0$$

for all α. Since $p(x, \omega)$ is strictly Pólya type 3, $p(x, \omega_i^0) - p(x, \omega_{i+1}^0)$ has at most one zero; by (9) it has at least one. In order that the relation (9) be preserved for all α, either $x_i(\alpha)$ increases and $x_{i+1}(\alpha)$ decreases, or $x_i(\alpha)$ decreases and $x_{i+1}(\alpha)$ increases, as α increases It is clear from (8) that the latter must hold.

It also follows from the variation diminishing properties of the density $p(x, \omega)$ [2] that

$$h_1(\omega) = \int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x)$$

is a monotone decreasing function of ω, and

$$h_n(\omega) = \int_{x_{n-1}}^{\infty} p(x, \omega)\, d\mu(x)$$

is a monotone increasing function of ω for any x_1 and x_{n-1} respectively.

Consider $x_i(\alpha)$ and $x_{i+1}(\alpha)$, which are defined by the equations $h_{i+1}(\omega_i^0) = \alpha = h_{i+1}(\omega_{i+1}^0)$, and $x'_{i+1}(\alpha)$ and $x'_{i+2}(\alpha)$, which are defined by $h_{i+2}(\omega_{i+1}^0) = \alpha = h_{i+2}(\omega_{i+2}^0)$. Then

LEMMA 3. *For all α, $x_i(\alpha) < x'_{i+1}(\alpha)$ and $x_{i+1}(\alpha) < x'_{i+2}(\alpha)$, $i = 1, \cdots, n-3$.*

Proof. Let

$$I_{[a,b]} = \begin{cases} 1, & a \leqq x \leqq b, \\ 0, & \text{otherwise.} \end{cases}$$

Suppose $x_i(\alpha) \geqq x'_{i+1}(\alpha)$. Then $I_{[x_i, x_{i+1}]} - I_{[x'_{i+1}, x'_{i+2}]}$ is always of one sign or at worse changes sign from $-$ to $+$. But

$$(10) \qquad \int_{-\infty}^{\infty} [I_{[x_i,x_{i+1}]} - I_{[x'_{i+1},x'_{i+2}]}] p(x, \omega)\, d\mu(x) \begin{cases} > 0 & \text{for} \quad \omega < \omega_{i+1}^0 \\ = 0 & \text{for} \quad \omega = \omega_{i+1}^0 \\ < 0 & \text{for} \quad \omega > \omega_{i+1}^0 \end{cases}$$

which is an impossibility in that it changes sign in the wrong direction [2] so $x_i(\alpha) < x'_{i+1}(\alpha)$.

Suppose $x_{i+1}(\alpha) \geqq x'_{i+2}(\alpha)$. Then $I_{[x_i, x_{i+1}]} - I_{[x'_{i+1}, x'_{i+2}]}$ is always of one sign which contradicts (10).

As $\alpha \to 1$, $x_i(\alpha)$, $x'_{i+1}(\alpha) \to -\infty$ and $x_{i+1}(\alpha)$, $x'_{i+2}(\alpha) \to +\infty$ (or the ends of the spectrum of μ), and as $\alpha \to 0$, $x_i(\alpha) \to x_i^*$, $x_{i+1}(\alpha) \to x_i^*$, and $x'_{i+1}(\alpha) \to$

x_{i+1}^*, $x'_{i+2}(\alpha) \to x_{i+1}^*$. Lemma 5 below asserts that $x_i^* < x_{i+1}^*$ but first it is necessary to prove Lemma 4.

LEMMA 4. *$\partial p(x_i^*, \omega)/\partial\omega$ does not vanish at ω_i^0 or ω_{i+1}^0 but does vanish for some ω_i^* where $\omega_i^0 < \omega_i^* < \omega_{i+1}^0$, $i = 1, \cdots, n-2$.*

Proof. By the mean value theorem for some $\omega_i^*(\alpha) \,\varepsilon\, [\omega_i^0, \omega_{i+1}^0]$,

$$\frac{\partial}{\partial\omega}\int_{x_i(\alpha)}^{x_{i+1}(\alpha)} p(x, \omega)\, d\mu(x)\,\Bigg|_{\omega=\omega_i^*(\alpha)} = \int_{x_i(\alpha)}^{x_{i+1}(\alpha)} \frac{\partial}{\partial\omega} p(x, \omega_i^*(\alpha))\, d\mu(x) = 0$$

for every $\alpha \,\varepsilon\, [0, 1]$. As $\alpha \to 0$, $\omega_i^*(\alpha) \to \omega_i^*$; $\partial p(x_i^*, \omega_i^*)/\partial\omega = 0$. Suppose $\omega_i^* = \omega_i^0$. Then, $\partial p(x_i^*, \omega)/\partial\omega > 0$ for $\omega > \omega_i^0$ which implies that $p(x_i^*, \omega_i^0) < p(x_i^*, \omega_{i+1}^0)$. Since $p(x, \omega)$ is continuous in each variable, there exists $\epsilon > 0$ such that $p(x, \omega_i^0) < p(x, \omega_{i+1}^0)$ for all x satisfying $|x - x_i^*| < \epsilon$. But this implies that for sufficiently small α,

$$\int_{x_i(\alpha)}^{x_{i+1}(\alpha)} p(x, \omega_i^0)\, d\mu(x) < \int_{x_i(\alpha)}^{x_{i+1}(\alpha)} p(x, \omega_{i+1}^0)\, d\mu(x),$$

a contradiction of the definition of $x_i(\alpha)$ and $x_{i+1}(\alpha)$. Similarly, $\omega_i^* \neq \omega_{i+1}^0$. Thus, $\omega_i^* \,\varepsilon\, (\omega_i^0, \omega_{i+1}^0)$.

LEMMA 5. *$x_i^* < x_{i+1}^*$, $i = 1, \cdots, n-2$.*

Proof. By Lemma 3, $x_i^* \leqq x_{i+1}^*$. Suppose $x_i^* = x_{i+1}^*$. Then, $\partial p(x_i^*, \omega_i^*)/\partial\omega = \partial p(x_i^*, \omega_{i+1}^*)/\partial\omega = 0$, where $\omega_i^* \,\varepsilon\, (\omega_i^0, \omega_{i+1}^0)$, $\omega_{i+1}^* \,\varepsilon\, (\omega_{i+1}^0, \omega_{i+2}^0)$, which is impossible by assumption.

This lemma can now be utilized to construct decision procedures possessing the property of M.P. For the 2-action problem any monotone procedure (defined by a single number x_1) is unbiased. In the 3-action problem each monotone procedure (x_1, x_2) which satisfies $h_2(\omega_1^0) = \alpha = h_2(\omega_2^0)$ for some $\alpha \,\varepsilon\, [0, 1]$ is unbiased. This means the monotone M.P. procedures are a one parameter family since once x_1 is specified as possible, x_2 and α are determined. For $n = 4$ consider $x_1(\alpha_1)$, $x_2(\alpha_1)$ defined by $h_2(\omega_1^0) = \alpha_1 = h_2(\omega_2^0)$ and $x'_2(\alpha_2)$, $x'_3(\alpha_2)$ defined by $h_3(\omega_2^0) = \alpha_2 = h_3(\omega_3^0)$, where α_1 and α_2 are chosen small enough to insure that $x_2(\alpha_1) < x'_2(\alpha_2)$. By Lemma 5 this is possible. Increase α_1 and α_2 until $x_2(\alpha_1) = x'_2(\alpha_2)$. The monotone procedure defined by $(x_1(\alpha_1), x_2(\alpha_1), x'_3(\alpha_2))$ has the property of M.P. Again the monotone M.P. procedures form a one parameter family since any point $y \,\varepsilon\, (x_1^*, x_2^*)$ will determine α_1 and α_2 by the condition that $x_2(\alpha_1) = y = x'_2(\alpha_2)$.

For the case of 5 actions the same method of construction is employed and a one parameter family of monotone M.P. decision procedures is designated. Define

$$x_1(\alpha_1), x_2(\alpha_1) \text{ by } h_2(\omega_1^0) = \alpha_1 = h_2(\omega_2^0),$$

$$x'_2(\alpha_2), x'_3(\alpha_2) \text{ by } h_3(\omega_2^0) = \alpha_2 = h_3(\omega_3^0),$$

$$x''_3(\alpha_3), x''_4(\alpha_3) \text{ by } h_4(\omega_3^0) = \alpha_3 = h_4(\omega_4^0),$$

where α_1, α_2, α_3 are chosen so small that $x_2(\alpha_1) < x'_2(\alpha_2)$ and $x'_3(\alpha_2) < x''_3(\alpha_3)$. Increase α_1 and α_3 until $x_2(\alpha_1) = x'_2(\alpha_2)$ and $x'_3(\alpha_2) = x''_3(\alpha_3)$. The monotone procedure $(x_1(\alpha_1), x_2(\alpha_1), x'_3(\alpha_2), x''_4(\alpha_3))$ has the property of maximum probabilities.

The family has only one parameter since the point $y \varepsilon (x_1^*, x_2^*)$ determines α_1, α_2, and α_3 through the relation $x_2(\alpha_1) = y = x_2'(\alpha_2)$. (Note that some values of y in the interval may not be legitimate parameter points. This will happen when the condition $y = x_2'(\alpha_2)$ is satisfied by an α_2 for which $x_3'(\alpha_2) > x_3^*$.)

When $n = 6$, the reader may verify that this method of construction breaks down. The difficulty is that $x_i(\alpha)$ does not have to decrease at the same rate at which $x_{i+1}(\alpha)$ increases. It may not be possible to choose α_2 and α_3 such that $x_3'(\alpha_2) = x_3''(\alpha_3)$ and still have $x_1^* < x_2'(\alpha_2)$ and $x_4''(\alpha_3) < x_4^*$.

For the cases $n = 3$, 4, and 5, note what has been accomplished by introducing the principle of M.P. The statistician, instead of having to choose a procedure from the class of all monotone procedures which is defined by $n - 1$ parameters, has only to choose from a class of procedures defined by only one parameter, those monotone procedures which have the additional property of maximum probabilities.

If the unknown parameter occurs in the density in the form of a translation parameter, that is $p(\xi, \omega) = p(\xi - \omega)$, $d\mu(\xi) = d\xi$, and $p(\cdot)$ is a symmetric function with respect to the origin, then any monotone procedure φ^0 defined by the critical numbers $x_1 < x_2 < \cdots < x_{n-1}$ such that

$$\frac{x_i + x_{i+1}}{2} = \frac{\omega_i^0 + \omega_{i+1}^0}{2} \text{ for } i = 1, 2, \cdots, n - 2$$

satisfies the property of M.P. The proof of this statement is straightforward and is omitted.

3. Unbiasedness in the sense of Lehmann—A decision procedure $\varphi(x)$ is said to be unbiased (in the sense of Lehmann or risk unbiased) if

$$E_\theta[L(\omega, \varphi(x))] \geqq E_\theta[L(\theta), \varphi(x))] \quad (11)$$

for all ω and θ, where $E_\theta(\cdot)$ represents the expected value given that the state of nature is θ. By specializing the loss function $L(\omega, a)$, it can be readily verified that this general definition of unbiasedness reduces to some of the classical notions. For a full discussion of the significance of this concept, the reader is referred to [7].

We search in this analysis to discover when unbiased procedures exist within the class of monotone procedures for the case of multiaction problems. An effective method of explicit construction of such procedures would also be desirable. Unfortunately, in general unbiased procedure need not exist. However, Theorem 1 below provides an affirmative answer for a substantial class of loss functions satisfying assumptions (a) and (b).

It should be emphasized that in contrast to the principle of M.P., which also embodies a generalization of the notion of unbiasedness in testing hypotheses, the present extension involves the specific loss functions in a fundamental way.

$$L_j(\omega) = L_i, \quad \text{for all } \omega \text{ in } S_i = (\omega_{i-1}^0, \omega_i^0], \quad (a)$$

$$i = 1, \cdots, n, \quad j = 1, \cdots, n.$$

Let $L_{ij} - L_{i+1,j} = a_{i,j}$.

(b) $0 \geqq a_{i1} \geqq a_{i2} \geqq \cdots \geqq a_{ii}$ and $a_{ii} < 0$;

$$a_{i,i+1} \geqq a_{i,i+2} \geqq \cdots a_{i,n} \geqq 0 \text{ and } a_{i,i+1} > 0 \text{ for } i = 1, 2, \cdots, n-1.$$

$$\text{Let } b_{ij} = \begin{cases} -a_{ij} & j = 1, \cdots, i \\ a_{ij} & j = i+1, \cdots, n. \end{cases} \qquad i = 1, \cdots, n-1$$

For $j \leqq$ i, $k \geqq i + 2$, $\qquad i = 1, \cdots, n-1,$

(c)
$$\begin{vmatrix} b_{i,j} & b_{i,k} \\ b_{i+1,j} & b_{i+1,k} \end{vmatrix} \geqq 0.$$

Two important examples of decision problems whose loss functions satisfy conditions (b) and (c) are worth noting.

(I)
$$L_j(\omega) = c \mid i - j \mid \qquad \text{for } \omega \text{ in } S_i .$$

This case is referred to as the discrete absolute error loss function.

(II)
$$L_j(\omega) = \begin{cases} 0 & \omega \varepsilon S_j , \\ c & \omega \not\varepsilon S_j . \end{cases}$$

The second example corresponds to the case where one assigns a constant loss c for any error and zero loss for a correct decision.

The fact that, if it exists, the monotone unbiased procedure is unique lends greater significance to this principle.

Examples I and II above are special cases of loss structures having the form $L_{ij} = f(\mid i - j \mid) = L_{\mid i-j\mid}$. Loss structures of this general pattern possess considerable interest since many practical problems arise in which the incurred losses can be assumed to be proportional to the magnitude of the error and unrelated to the type of error. In the event that $L_{ij} = L_{\mid i-j\mid}$ (we say L_{ij} has a convolution form), condition (b) implies that $L_{\mid i-j\mid}$ is a concave function of $\mid i - j \mid$, i.e., $L_{r+1} \geqq \frac{1}{2}(L_r + L_{r+2})$, $r = 0, 1, \cdots, n - 2$. This is to say the loss increases concavely as the action actually taken diverges from the correct action. That concavity implies condition (b) is also true, so condition (b) is fully equivalent to the concavity of $L_{\mid i-j\mid}$ as a function of $\mid i - j \mid$. Moreover, condition (c) is automatically satisfied if $L_{\mid i-j\mid}$ is concave since $b_{ij} \geqq b_{i+1,j}$ for $j \leqq i$ and $b_{i,k} \leqq b_{i+1,k}$ for $k \geqq i + 2$. Therefore, for this convolution case, the hypotheses of Theorem 1 are equivalent to the statement that $L_{\mid i-j\mid}$ is a concave function of its argument.

It should be noted that condition (c) is not the same as condition (II) of [3]. However, in the important case $L_{ij} = L_{\mid i-j\mid}$, the two conditions are equivalent since the two b_{ij} matrices are identical. Consequently, when the loss function $L_{ij} = L_{\mid i-j\mid}$ is concave, all non-degenerate monotone procedures are admissible.

In particular, the unique unbiased procedure guaranteed by Theorem 1 which is also shown to be non-degenerate (Corollary 4) is necessarily admissible in the case where L_{ij} is of the convolution form.

(The proof of Theorem 2 of [3] is easily seen to apply in the case of loss functions of convolution form satisfying (b) and (c), above.)

The principle theorem concerning unbiased procedures is the following:

THEOREM 1. *If assumptions* (a), (b), *and* (c) *are satisfied, then there exists a unique monotone procedure which is unbiased in the sense of Lehmann.*

To avoid inessential tedious details we assume that $p(x, \omega)$ is strictly Pólya type 2, and μ is a continuous measure whose spectrum is an interval. The analogous results when the assumption on μ is relaxed are immediate.

The proof of Theorem 1 is more elaborate and will be presented in Sec. 4. We dwell in this section on the important special case of (I) where the proofs are considerably simpler and for which some additional results are obtained (Theorem 2).

Proof of Theorem 1 *for the special case* (I). For a monotone procedure $(x_1, \cdots, x_{n-1})$ define

$$A_1(\omega) = c\int_{x_1}^{x_2} p(x, \omega)\, d\mu(x) + 2c\int_{x_2}^{x_3} p(x, \omega)\, d\mu(x) + \cdots + (n-1)c\int_{x_{n-1}}^{\infty} p(x, \omega)\, d\mu(x)$$

$$A_2(\omega) = c\int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x) + c\int_{x_2}^{x_3} p(x, \omega)\, d\mu(x) + \cdots + (n-2)c\int_{x_{n-1}}^{\infty} p(x, \omega)\, d\mu(x)$$

$$\vdots$$

$$A_n(\omega) = (n-1)c\int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x) + (n-2)c\int_{x_1}^{x_2} p(x, \omega)\, d\mu(x) + \cdots + c\int_{x_{n-2}}^{x_{n-1}} p(x, \omega)\, d\mu(x).$$

For $\omega \varepsilon S_i$, $i = 1, \cdots, n$, $\rho(\omega, \varphi) = A_i(\omega)$. Define

$$B_i(\omega) = A_i(\omega) - A_{i+1}(\omega), \qquad i = 1, \cdots, n-1.$$

It is immediate that

$$B_i(\omega) = -c\int_{-\infty}^{x_i} p(x, \omega)\, d\mu(x) + c\int_{x_i}^{\infty} p(x, \omega)\, d\mu(x), \qquad i = 1, \cdots, n-1.$$

In order that the monotone procedure be unbiased it is necessary and sufficient that $B_j(\omega) \geqq 0, j = 1, \cdots, i-1$; $B_j(\omega) \leqq 0, j = i, \cdots, n-1$ for $\omega \varepsilon S_i$, $i = 1, \cdots, n-1$. Choose the unique $x_1 = x_1^0$ which satisfies $B_1(\omega_1^0) = 0$. Then $B_1(\omega) \ (\lesseqgtr)\ 0$ for all $\omega \ (\lesseqgtr)\ \omega_1^0$. Since $x_i \geqq x_1^0$ for $i = 2, \cdots, n-1$, $B_i(\omega) < 0$ for $\omega < \omega_1^0$, $i = 2, \cdots, n-1$. Unbiasedness further requires that for $\omega \varepsilon S_2$ $B_1(\omega) \geqq 0$ and $B_i(\omega) \leqq 0$ for $i = 2, \cdots, n-1$. Determine the unique $x_2 = x_2^0$ such that $B_2(\omega_2^0) = 0$. $x_2^0 > x_1^0$ since $\omega_2^0 > \omega_1^0$, and $B_i(\omega) < 0$ for $\omega \varepsilon S_2$ and $i =$

$2, \cdots, n-1$. The continuation of this construction will produce the unique monotone unbiased procedure $(x_1^0, \cdots, x_{n-1}^0)$.

For the special loss function under consideration this unique unbiased procedure is uniformly most powerful within the class of all unbiased procedures. This is the substance of the following theorem which is a special case of Theorem 2 of [8]. The proof is included by merit of its simplicity and because it also illustrates on a small scale some of the ideas necessary in carrying out the arguments of Theorem 1.

THEOREM 2. *If* $L_i(\omega) = c \mid i - j \mid$ *for* ω *in* S_j, *then any unbiased procedure* $\varphi = (\varphi_1, \cdots, \varphi_n)$ *is everywhere improved upon by the unique monotone unbiased procedure, except possibly at* $\omega_1^0, \cdots, \omega_{n-1}^0$.

Proof. By definition,

$$B_1(\omega) = A_1(\omega) - A_2(\omega) = -c \int_{-\infty}^{\infty} \varphi_1(x) p(x, \omega)\, d\mu(x) + c \int_{-\infty}^{\infty} (1 - \varphi_1(x)) p(x, \omega)\, d\mu(x)$$

$$= c - 2c \int_{-\infty}^{\infty} \varphi_1(x) p(x, \omega)\, d\mu(x),$$

and for $k = 2, \cdots, n-1$,

$$B_k(\omega) = A_k(\omega) - A_{k+1}(\omega) = c - 2c \int_{-\infty}^{\infty} [\varphi_1(x) + \cdots + \varphi_k(x)] p(x, \omega)\, d\mu(x).$$

Consider any other decision procedure φ^* which is not necessarily unbiased. For $k = 1, \cdots, n-1$,

$$B_k^{\varphi}(\omega) - B_k^{\varphi^*}(\omega) = 2c \int_{-\infty}^{\infty} [(\varphi_1^*(x) + \cdots + \varphi_k^*(x)) - (\varphi_1(x) + \cdots + \varphi_k(x)] p(x, \omega)\, d\mu(x).$$

If φ^* is the monotone procedure constructed so that it improves upon φ according to Lemma 4 of [4], then φ^* satisfies

$$B_k^{\varphi}(\omega) - B_k^{\varphi^*}(\omega) \begin{cases} \geqq 0 & \text{for } \omega \leqq \omega_k^0 \\ \leqq 0 & \text{for } \omega \geqq \omega_k^0 \end{cases}$$

for $k = 1, \cdots, n-1$. But $B_k^{\varphi}(\omega_k^0) = 0$. Therefore, $B_k^{\varphi^*}(\omega_k^0) = 0$ which implies that φ^* is unbiased. Since there is only one monotone unbiased procedure, φ^* must be identical with the φ^0 of Theorem 1.

The limiting case of an n-action problem as $n \to +\infty$ is an estimation problem. Suppose that for the problem under consideration the limit is taken in such a manner that as $n \to \infty$, $\omega_1^0 \to -\infty$, $\omega_{n-1}^0 \to +\infty$, $\mid \omega_i^0 - \omega_{i-1}^0 \mid \to 0$, $i = 2, \cdots, n-1$, and $L_n(\omega, i_n(a)) \to c \mid a - \omega \mid$, where $i_n(a)$ is defined by $a \,\varepsilon\, S_{i_n(a)}$. The resulting problem is an estimation problem with absolute error loss function. It is easily verified that the estimate $\delta(x)$, which is the limit of the unique monotone

procedures which are unbiased in the sense of Lehmann, is defined by the relation

$$\int_{-\infty}^{x} p(y, \delta(x))\, d\mu(y) = \int_{x}^{\infty} p(y, \delta(x))\, d\mu(y). \tag{12}$$

This, of course, states the well-known fact that the median unbiased estimate of θ is the function $\delta(x)$ which satisfies (12) when x is observed.

4. *Proof of Theorem* 1. For purposes of clarity the proof of the theorem is divided into a series of separate steps. First, we introduce the relevant quantities entering into the analysis. For a procedure $\varphi = (\varphi_1, \varphi_2, \cdots, \varphi_n)$, let

$$A_i^{\varphi}(\omega) = L_{i1} \int_{-\infty}^{\infty} \varphi_1(x) p(x, \omega)\, d\mu(x) + \cdots + L_{in} \int_{-\infty}^{\infty} \varphi_n(x) p(x, \omega) d\mu(x)$$

for $i = 1, \cdots, n$. When ω ranges over S_i the function $A_i^{\varphi}(\omega)$ coincides with $\rho(\omega, \varphi)$, the expected risk. Also for $i = 1, 2, \cdots, n-1$, we define

$$\begin{aligned} B_i^{\varphi}(\omega) &= A_i^{\varphi}(\omega) - A_{i+1}^{\varphi}(\omega) \\ &= a_{i1} \int_{-\infty}^{\infty} \varphi_1(x) p(x, \omega)\, d\mu(x) + \cdots + a_{in} \int_{-\infty}^{\infty} \varphi_n(x) p(x, \omega)\, d\mu(x) \\ &= -b_{i1} \int_{-\infty}^{\infty} \varphi_1(x) p(x, \omega)\, d\mu(x) - \cdots - b_{ii} \int_{-\infty}^{\infty} \varphi_i(x) p(x, \omega)\, d\mu(x) \\ &\quad + b_{i,i+1} \int_{-\infty}^{\infty} \varphi_{i+1}(x) p(x, \omega)\, d\mu(x) + \cdots + b_{in} \int_{-\infty}^{\infty} \varphi_n(x) p(x, \omega)\, d\mu(x). \end{aligned} \tag{13}$$

If a decision procedure φ satisfies the system of inequalities

$$\begin{aligned} B_k^{\varphi}(\omega) &\geqq 0 \qquad 1 \leqq k \leqq i-1 \\ B_k^{\varphi}(\omega) &\leqq 0 \qquad i \leqq k \leqq n-1 \end{aligned} \qquad \text{and } \omega \text{ in } S_i, \tag{14}$$

then φ is clearly unbiased in the sense of Lehmann. In general, the converse is not valid. However, it is true that for monotone procedures the property of unbiasednesss implies that this system of inequalities is satisfied. The inequalities are fulfilled for a monotone procedure $\varphi = (x_1, x_2, \cdots, x_{n-1})$ if and only if

$$B_i^{\varphi}(\omega_i^0) = 0, \qquad i = 1, 2, \cdots, n-1. \tag{15}$$

In fact, the variation diminishing properties of the density $p(x, \omega)$ imply that $B_i^{\varphi}(\omega) < 0$ for $\omega < \omega_i^0$ and $B_i^{\varphi}(\omega) > 0$ for $\omega > \omega_i^0$ which in turn are equivalent to the system of inequalities (14). Our problem reduces to the demonstration of the existence and uniqueness of a set of values $x = (x_1, x_2, \cdots, x_{n-1})$ where $x_1 \leqq x_2 \leqq \cdots \leqq x_{n-1}$ which are a solution to the system of non-linear equations:

$$\begin{aligned} B_i^{x}(\omega_i^0) &= -b_{i1} \int_{-\infty}^{x_1} p(\xi, \omega_i^0)\, d\mu(\xi) - \cdots - b_{ii} \int_{x_{i-1}}^{x_i} p(\xi, \omega_i^0)\, d\mu(\xi) \\ &\quad + b_{i,i+1} \int_{x_i}^{x_{i+1}} p(\xi, \omega_i^0)\, d\mu(\xi) + \cdots + b_{i,n} \int_{x_{n-1}}^{\infty} p(\xi, \omega_i^0)\, d\mu(\xi) = 0. \end{aligned} \tag{16}$$

Turning to this task we start by showing that the mapping $x \to y$ whic defined coordinate-wise by $y_i = B_i^x(\omega_i^0)$, $i = 1, \cdots, n-1$, and which maps $n-1$ dimensional simplex of all $n-1$ tuples $x = (x_1, x_2, \cdots, x_{n-1})$ satisf $x_1 \leqq x_2 \leqq \cdots \leqq x_{n-1}$ into Euclidean $n-1$ dimensional space (E^{n-1}) is a to-one mapping. Precisely:

LEMMA 6. *The mapping $y_i = B_i^x(\omega_i^0)$, $i = 1, \cdots, n-1$, defined on the set o monotone procedures by means of the formulas* (16) *with image in E^{n-1} space one-to-one transformation.*

Proof (*by contradiction*). Suppose there exist two different monotone proced $\varphi \sim x = (x_1, x_2, \cdots, x_{n-1})$ and $\varphi' \sim x' = (x_1', x_2', \cdots, x_{n-1}')$ with the prop that $B_i^{x'}(\omega_i^0) - B_i^x(\omega_i^0) = 0$ for $i = 1, \cdots, n-1$. Without loss of genera assume $x_1' \geqq x_1$. $B_i^{x'}(\omega_i^0) - B_i^x(\omega_i^0) = 0$, $i = 1, \cdots, n-1$, yields the sys of equations

$$0 = -(b_{11} + b_{12}) \int_{x_1}^{x_1'} p(x, \omega_1^0)\, d\mu(x) + (b_{12} - b_{13}) \int_{x_2}^{x_2'} p(x, \omega_1^0)\, d\mu(x)$$
$$+ \cdots + (b_{1,n-1} - b_{1n}) \int_{x_{n-1}}^{x_{n-1}'} p(x, \omega_1^0)\, d\mu$$
$$0 = (b_{22} - b_{21}) \int_{x_1}^{x_1'} p(x, \omega_2^0)\, d\mu(x) - (b_{22} + b_{23}) \int_{x_2}^{x_2'} p(x, \omega_2^0)\, d\mu(x)$$
$$+ \cdots + (b_{2,n-1} - b_{2,n}) \int_{x_{n-1}}^{x_{n-1}'} p(x, \omega_2^0)\, d\mu$$
$$\vdots$$
$$0 = (b_{n-1,2} - b_{n-1,1}) \int_{x_1}^{x_1'} p(x, \omega_{n-1}^0)\, d\mu(x)$$
$$+ \cdots + (b_{n-1,n-1} - b_{n-1,n-2}) \int_{x_{n-2}}^{x_{n-2}'} p(x, \omega_{n-1}^0)\, d\mu$$
$$- (b_{n-1,n-1} + b_{n-1,n}) \int_{x_{n-1}}^{x_{n-1}'} p(x, \omega_{n-1}^0)\, d\mu(x).$$

Since $(b_{11} + b_{12}) > (b_{12} - b_{13} + \cdots + (b_{1,n-1} - b_{1,n})$, it follows that there ex a k, $1 < k \leqq n-1$, such that

$$\int_{x_l}^{x_l'} p(x, \omega_1^0)\, d\mu(x) < \int_{x_k}^{x_k'} p(x, \omega_1^0)\, d\mu(x)$$

for $1 \leqq l < k$. If k is not unique, choose the largest k which satisfies this pr erty. Consider the kth equation. For $1 \leqq l < k$,

$$\int_{x_l}^{x_l'} p(x, \omega_k^0)\, d\mu(x) < \int_{x_k}^{x_k'} p(x, \omega_k^0)\, d\mu(x)$$

by the fundamental change of sign property for strictly Pólya type 2 densi since $x_l \leqq x_k$ and $x_l' < x_k'$. But $(b_{kk} + b_{k,k+1}) \geqq (b_{k2} - b_{k1}) + \cdots + (b_{kk} - b_{k,}$

$+ (b_{k,k+1} - b_{k,k+2}) + \cdots + (b_{k,n-1} - b_{k,n})$. Therefore on examination of the kth equation, if $k < n - 1$, there exists an $h > k$ for which

$$\int_{x_l}^{x_l'} p(x, \omega_k^0)\, d\mu(x) < \int_{x_h}^{x_h'} p(x, \omega_k^0)\, d\mu(x)$$

for $1 \leqq l < h$. If h is not unique, choose the largest h.

Continue this argument until at the last step it has been established that

$$\int_{x_l}^{x_l'} p(x, \omega_{n-1}^0)\, d\mu(x) < \int_{x_{n-1}}^{x_{n-1}'} p(x, \omega_{n-1}^0)\, d\mu(x)$$

for $1 \leqq l < n - 1$. But this contradicts the fact that $B_{n-1}^{x'}(\omega_{n-1}^0) - B_{n-1}^{x}(\omega_{n-1}^0) = 0$ since $(b_{n-1,n-1} + b_{n-1,n}) > (b_{n-1,2} - b_{n-1,1}) + \cdots + (b_{n-1,n-1} - b_{n-1,n-2})$.

Corollary 1. *There exists at most one monotone unbiased procedure.*

The proof is immediate. We shall need the following slight extension of Lemma 6.

Corollary 2. If φ is the monotone procedure $x = (x_1, x_2, \cdots, x_{n-1})$ and $\varphi' \sim x' = (x_1', x_2', \cdots, x_{n-1}')$ with $x_{n-1}' \geqq x_{n-1}$ and $B_i^{x'}(\omega_i^0) - B_i^{x}(\omega_i^0) \geqq 0$ for $i = 1, 2, \cdots, n - 1$, then $x_i' = x_i$ for $i = 1, 2, \cdots, n - 1$.

The proof of Corollary 2 is essentially a paraphrase of that of Lemma 6. We sketch the details. Let k be the first index where $x_k' \geqq x_k$ $(k \leqq n - 1)$. By examining the kth relation $B_k^{x'}(\omega_k^0) - B_k^{x}(\omega_k^0) \geqq 0$ as in the proof of the lemma, when $k < n - 1$, we may find a larger index $h > k$ such that for $i < h$,

$$\int_{x_i}^{x_i'} p(\xi, \omega_k^0)\, d\mu(\xi) < \int_{x_h}^{x_h'} p(\xi, \omega_k^0)\, d\mu(\xi).$$

From the variation diminishing properties of $p(\xi, \omega)$ we may conclude that for $i < h$,

$$\int_{x_i}^{x_i'} p(\xi, \omega_h^0)\, d\mu(\xi) < \int_{x_h}^{x_h'} p(\xi, \omega_h^0)\, d\mu(\xi).$$

On continued inspection of the hth relation, we find a larger index until we reach the $(n - 1)^{\text{th}}$ index with the property that

$$\int_{x_i}^{x_i'} p(\xi, \omega_{n-1}^0)\, d\mu(\xi) < \int_{x_{n-1}}^{x_{n-1}'} p(\xi, \omega_{n-1}^0)\, d\mu(\xi), \qquad i = 1, 2, \cdots, n - 2.$$

The last inequality

$$B_{n-1}^{x'}(\omega_{n-1}^0) - B_{n-1}^{x}(\omega_{n-1}^0) \geqq 0$$

is evidently contradicted.

One final extension in the same direction is the following:

Corollary 3. *If $\varphi \sim (x_1, x_2, \cdots, x_{n-1})$ and $\varphi' \sim (x_1', x_2', \cdots, x_{k-1}', \gamma, \cdots, \gamma)$*

are two monotone procedures such that $B_i^{\varphi'}(\omega_i^0) - B_i^{\varphi}(\omega_i^0) \geqq 0$ *for* $i = 1, \cdots, k - 1$ *and* $\gamma \geqq x_{n-1}$, *then*

$$B_k^{\varphi'}(\omega_k^0) - B_k^{\varphi}(\omega_k^0) \leqq 0.$$

The proof follows the same line of reasoning as the preceding.

In view of Corollary 1 it remains to prove the existence part of Theorem 1. We require the following lemma.

LEMMA 7. *Let the* $2 \times m$ *matrix* (e_{ij}), $i = 1, 2, j = 1, \cdots, m$, *consist of non-negative elements, and let* $\lambda_1, \cdots, \lambda_m$ *be non-negative constants. Let condition* (E) *be satisfied:*

$$\text{(E)} \qquad \begin{vmatrix} e_{1j} & e_{1k} \\ e_{2j} & e_{2k} \end{vmatrix} \geqq 0$$

for $1 \leqq j \leqq l$, $l + 2 \leqq k \leqq m$. *If* $0 < e_{11}\lambda_1 + \cdots + e_{1l}\lambda_l \leqq e_{1,l+2}\lambda_{l+2} + \cdots + e_{1,m}\lambda_m$, *then* $e_{21}\lambda_1 + \cdots + e_{2l}\lambda_l \leqq e_{2,l+2}\lambda_{l+2} + \cdots + e_{2m}\lambda_m$.

Proof. By (E), $\sum_{j=1}^{l} (e_{2k}e_{1j} - e_{1k}e_{2j})\lambda_j \geqq 0$ for $k \geqq l + 2$. Therefore, $e_{2k} \sum_{j=1}^{l} e_{1j}\lambda_j \geqq e_{1k} \sum_{j=1}^{l} e_{2j}\lambda_j$, and $\sum_{k=l+2}^{m} e_{2k}\lambda_k \cdot \sum_{j=1}^{l} e_{1j}\lambda_j \geqq \sum_{k=l+2}^{m} e_{1k}\lambda_k \cdot \sum_{j=1}^{l} e_{2j}\lambda_j$. For $0 < \sum_{j=1}^{l} e_{1j}\lambda_j \leqq \sum_{k=l+2}^{m} e_{1k}\lambda_k$, $\sum_{k=l+2}^{m} e_{2k}\lambda_k \geqq \sum_{j=1}^{l} e_{2j}\lambda_j$.

Proof of existence. It suffices to show there exists a monotone φ for which $B_i^{\varphi}(\omega_i^0) = 0, i = 1, \cdots, n - 1$. This holds trivially for $n = 2$. Suppose it is true for the case of n actions. The argument is inductive. For $n + 1$ actions and a monotone procedure, let

$$B_i^{\varphi}(\omega) = -b_{i1} \int_{-\infty}^{x_1} p(x, \omega)\, d\mu(x) - \cdots - b_{ii} \int_{x_{i-1}}^{x_i} p(x, \omega)\, d\mu(x)$$
$$+ b_{i,i+1} \int_{x_i}^{x_{i+1}} p(x, \omega)\, d\mu(x) + \cdots + b_{i,n+1} \int_{x_n}^{\infty} p(x, \omega)\, d\mu(x)$$

for $i = 1, \cdots, n$.

(1) Choose $x_n = \infty$. The conditions (a), (b), and (c) are fulfilled so by the induction hypothesis there exists a solution $\varphi^{\infty} \sim (x_1^{\infty}, \cdots, x_{n-1}^{\infty}, \infty)$ of the system of equations $B_i^{\varphi}(\omega_i^0) = 0, i = 1, \cdots, n - 1$. For this solution obviously $B_n^{\infty}(\omega_n^0) \leqq 0$.

(2) Choose $x_{n-1} = x_n$. By the induction hypothesis there exists a solution $\varphi^{x^0} \sim (x_1^0, \cdots, x_{n-1}^0 = x^0, x_n^0 = x^0)$ of $B_i^{\varphi}(\omega_i^0) = 0, i = 1, \cdots, n - 1$. Since $B_{n-1}^{x^0}(\omega_{n-1}^0) = 0$, the variation diminishing properties of densities possessing a strict monotone likelihood ratio lead to the conclusion that $B_{n-1}^{x^0}(\omega_n^0) \geqq 0$. If $B_{n-1}^{x^0}(\omega_n^0) = 0$, then it follows that $x^0 = -\infty$ which in turn implies that $B_n^{x^0}(\omega_n^0) > 0$. On the other hand, if $B_{n-1}^{x^0}(\omega_n^0) > 0$, let $l = n - 1$, $m = n$, $e_{1j} = b_{n-1,j}$ for $j = 1, \cdots, n - 1$, $e_{1n} = b_{n-1,n+1}$, $e_{2j} = b_{n,j}$ for $j = 1, \cdots, n - 1$ and $e_{2n} = b_{n,n+1}$ in Lemma 7. Then, by Lemma 7 $B_{n-1}^{x^0}(\omega_n^0) > 0$ implies $B_n^{x^0}(\omega_n^0) \geqq 0$.

It has been shown thus far that there exists a strategy $(x_1^{\infty}, \cdots, x_{n-1}^{\infty}, \infty)$ such that $B_i^{\varphi}(\omega_i^0) = 0$, $i = 1, \cdots, n - 1$, and $B_n^{\varphi}(\omega_n^0) \leqq 0$ and a strategy $(x_1^0, \cdots, x_{n-1}^0 = x^0, x_n^0 = x^0)$ such that $B_i^{\varphi}(\omega_i^0) = 0$, $i = 1, \cdots, n - 1$ and $B_n^{\varphi}(\omega_n^0) \geqq 0$. If it can be shown that for every x_n satisfying $x^0 < x_n < \infty$ there

exists a solution $(x_1, \cdots, x_{n-1})$ to $B_i^{\varphi}(\omega_i^0) = 0$ $i = 1, \cdots, n-1$, then by continuity a solution exists satisfying $B_i^{\varphi}(\omega_i^0) = 0$, $i = 1, \cdots, n$; the continuity of the solution as a function of x_n being a simple consequence of Lemma 6.

The proof that for every z, $x^0 < z < \infty$, there exist $(x_1(z), \cdots, x_{n-1}(z))$ such that $\varphi \sim (x_1(z), \cdots, x_{n-1}(z), z)$ satisfies $B_i^{\varphi}(\omega_i^0) = 0$, $i = 1, \cdots, n-1$, proceeds in a stepwise manner.

(3) Let $x_1 \leqq x_2 = \cdots = x_{n-1} = z$.

(a) Choose $x_1 = z$. Since $b_{12} \geqq b_{13} \geqq \cdots \geqq b_{1,n+1}$,

$$-b_{11}\int_{-\infty}^{x_1^0} p(x, \omega_1^0)\,d\mu(x) + b_{1,n+1}\int_{x_1^0}^{\infty} p(x, \omega_1^0)\,d\mu(x) \leqq 0$$

which implies

$$-b_{11}\int_{-\infty}^{z} p(x, \omega_1^0)\,d\mu(x) + b_{1,n+1}\int_{z}^{\infty} p(x, \omega_1^0)\,d\mu(x) \leqq 0$$

since $x_1^0 \leqq x_n^0 < z$.

(b) Choose $x_1 = -\infty$.

$$b_{12}\int_{-\infty}^{z} p(x, \omega_1^0)\,d\mu(x) + b_{1,n+1}\int_{z}^{\infty} p(x, \omega_1^0)\,d\mu(x) \geqq 0.$$

(c) Thus by continuity there must exist an $x_1^1 = x_1^1(z)$ which satisfies

$$-b_{11}\int_{-\infty}^{x_1^1} p(x, \omega_1^0)\,d\mu(x) + b_{12}\int_{x_1^1}^{z} p(x, \omega_1^0)\,d\mu(x) + b_{1,n+1}\int_{z}^{\infty} p(x, \omega_1^0)\,d\mu(x) = 0.$$

(4) Let $x_1 \leqq x_2 \leqq x_3 = \cdots = x_{n-1} = z$. Consider the two expressions

$$c_1(\omega; x_1, x_2) = -b_{11}\int_{-\infty}^{x_1} p(x, \omega)\,d\mu(x) + b_{12}\int_{x_1}^{x_2} p(x, \omega)\,d\mu(x)$$
$$+ b_{13}\int_{x_2}^{z} p(x, \omega)\,d\mu(x) + b_{1,n+1}\int_{z}^{\infty} p(x, \omega)\,d\mu(x),$$

$$c_2(\omega; x_1, x_2) = -b_{21}\int_{-\infty}^{x_1} p(x, \omega)\,d\mu(x) - b_{22}\int_{x_1}^{x_2} p(x, \omega)\,d\mu(x)$$
$$+ b_{23}\int_{x_2}^{z} p(x, \omega)\,d\mu(x) + b_{2,n+1}\int_{z}^{\infty} p(x, \omega)\,d\mu(x).$$

Of course $c_j(\omega, x_1, x_2) = B_j^{\varphi}(\omega)$, $j = 1, 2$, for the special procedure $\varphi \sim (x_1, x_2, z, \cdots z)$. Our immediate object now is to show that x_1 and x_2 exist satisfying $(x_1 \leqq x_2 \leqq z)$ such that $c_1(\omega_1^0; x_1, x_2) = 0$ and $c_2(\omega_2^0; x_1, x_2) = 0$.

(a) Choose $x_2 = z$. By (3) above there exists an $x_1'(z)$ for which

$$c_1(\omega_1^0; x_1'(z), z) = 0.$$

We assert that $c_2(\omega_2^0; x_1'(z), z) \leqq 0$. Comparing for $i = 1, 2$ $B_i^{\varphi^0}(\omega_i^0)$ and $B_i^{\varphi}(\omega_i^0)$ where $\varphi^0 \sim (x_1^0, x_2^0, \cdots, x_{n-1}^0 = x^0, x_n^0 = x^0)$ of (2) above and

$\varphi \sim (x_1'(z), z, z, \cdots, z)$ with $z > x^0$, we see the conditions of Corollary 3 are met and therefore we may conclude $c_2(\omega_2^0, x_1'(z), z) \leqq 0$ as stated.

(b) Choose $x_1 = x_2$; then $c_1(\omega_1^0; -\infty, -\infty) \geqq 0$ and $c_1(\omega_1^0; z, z) \leqq 0$ by (3a). Thus there exists a $u = x_1 = x_2$ such that $c_1(\omega_1^0; u, u) = 0$ which implies $c_1(\omega_2^0; u, u) \geqq 0$ If $c_1(\omega_2^0; u, u) = 0$, then $u = -\infty$ which in turn implies $c_2(\omega_2^0; u, u) \geqq 0$. If in the other circumstance $c_1(\omega_2^0; u, u) > 0$, then by Lemma 7 we infer that $c_2(\omega_2^0; u, u) \geqq 0$.

(c) We next prove that there exists an $x_1^* = x_1^*(y)$ such that $c_1(\omega_1^0; x_1^*, y) = 0$ for every $u < y < z$. (This is like the larger problem we are trying to solve for the special case when $n = 2$. The quantity z plays the role of ∞ and u adopts the role of z.) When $x_1 = y$, $c_1(\omega_1^0; y, y) < 0$ because $c_1(\omega_1^0; u, u) = 0$ and $y > u$. Obviously $c_1(\omega_1^0; -\infty, y) > 0$. By continuity there exists an x_1^* such that $c_1(\omega_1^0; x_1^*, y) = 0$.

Since $c_1(\omega_1^0; x_1, y) = 0$ has a solution x_1^* for every y in the interval $[u, z]$ and $c_2(\omega_2^0; x_1'(z), z) \leqq 0$, $c_2(\omega_2^0; u, u) \geqq 0$, by continuity there must exist an $x_2^2(z) = y \in [u, z]$ and $x_1^2(z)$ such that $c_1(\omega_1^0; x_1^2, x_2^2) = c_2(\omega_2^0; x_1^2, x_2^2) = 0$.

(5) Let $x_1 \leqq x_2 \leqq x_3 \leqq x_4 = \cdots = z$. Consider the three expressions

$$D_i(\omega; x_1, x_2, x_3) = -\sum_{j=1}^{i} b_{ij} \int_{x_{j-1}}^{x_j} p(x, \omega)\, d\mu(x) + \sum_{j=i+1}^{3} b_{ij} \int_{x_{j-1}}^{x_j} p(x, \omega)\, d\mu(x) + b_{i4} \int_{x_3}^{z} p(x, \omega)\, d\mu(x) + b_{i,n+1} \int_{z}^{\infty} p(x, \omega)\, d\mu(x),$$

$i = 1, 2, 3$, where $x_0 = -\infty$. Of course $D_i(\omega; x_1, x_2, x_3) = B_i^\varphi(\omega)$ where $\varphi \sim (x_1, x_2, x_3, z, z, \cdots, z)$. The next step is to try to solve $D_i(\omega_i^0; x_1, x_2, x_3) = 0$, $i = 1, 2, 3$.

(a) Choose $x_3 = z$. By (4) above there exists a couple $(x_1^2(z), x_2^2(z))$ such that $D_1(\omega_1^0; x_1^2(z), x_2^2(z), z) = D_2(\omega_2^0; x_1^2(z), x_2^2(z), z) = 0$. Corollary 3 may be applied and we find that on comparison with the relations $B_i^{\varphi^0}(\omega_i) = 0$, $i = 1, 2, 3$, for $\varphi^0 \sim (x_1^0, x_2^0, \cdots, x_n^0)$ of (2), $D_3(\omega_3^0; x_1^2(z), x_2^2(z), z) \leqq 0$.

(b) Choose $x_2 = x_3$. By (4) there exists a solution $(\bar{x}_1(w), w)$ where $x_2 = x_3 = w$ to the equations $D_1(\omega_1^0; x_1, x_2, x_2) = 0$, $D_2(\omega_2^0; x_1, x_2, x_2) = 0$. $D_3(\omega_3^0; \bar{x}_1, w, w) \geqq 0$ is a consequence of Lemma 7.

(c) There exists a couple $(x_1^{**}(y), x_2^{**}(y))$ such that

$$D_1(\omega_1^0; x_1^{**}(y), x_2^{**}(y), y) = D_2(\omega_2^0; x_1^{**}(y), x_2^{**}(y), y) = 0$$

for every $y \in [w, z]$.

The proof of this step requires a repetition of the previous arguments as carried out for the function c_i with y taking the role of ∞. To this end, we establish

(c.1) Choose $x_1 = x_2$. For $x_1 = -\infty$, $D_1(\omega_1^0; -\infty, -\infty, y) \geqq 0$. For $x_1 = y$, $D_1(\omega_1^0; y, y, y) \leqq 0$ since $D_1(\omega_1^0; \bar{x}_1(w), w, w) = 0$ implies $D_1(\omega_1^0; w, w, w) \leqq 0$ and $y > w$. Therefore, there exists $x_1 = v$ such that $D_1(\omega_1^0; v, v, y) = 0$. It can be shown by applying Lemma 7 that $D_2(\omega_2^0; v, v, y) \geqq 0$.

(c.2) Choose $x_2 = y$. $D_1(\omega_1^0; -\infty, y, y) \geqq 0$ and $D_1(\omega_1^0; y, y, y) \leqq 0$. Thus, there exists $x_1(y)$ such that $D_1(\omega_1^0; x_1(y), y, y) = 0$. $D_2(\omega_2^0; x_1(y), y, y) \leqq 0$. The last inequality may be deduced from Corollary 3 by comparing the procedures $\varphi' \sim (x_1(y), y, y, z, z, \cdots, z)$ and $\varphi \sim (\bar{x}_1(w), w, w, z, z, \cdots, z)$.

In fact, suppose the inequality $D_2(\omega_2^0; x_1(y), y, y) \leqq 0$ is violated. Consider the solution $(\bar{x}_1(w), w, w)$ to the system of equations

$$D_1(\omega_1^0; x_1, x_2, x_2) = D_2(\omega_2^0; x_1, x_2, x_2) = 0.$$

$$D_1(\omega_1^0; x_1(y), y, y) - D_1(\omega_1^0; \bar{x}_1(w), w, w)$$
$$= -(b_{11} + b_{12}) \int_{\bar{x}_1(w)}^{x_1(y)} p(x, \omega_1^0)\, d\mu(x) + (b_{12} - b_{14}) \int_w^y p(x, \omega_1^0)\, d\mu(x) = 0, \tag{17}$$

$$D_2(\omega_2^0; x_1(y), y, y) - D_2(\omega_2^0; \bar{x}_1(w), w, w)$$
$$= (b_{22} - b_{21}) \int_{\bar{x}_1(w)}^{x_1(y)} p(x, \omega_2^0)\, d\mu(x) - (b_{22} + b_{24}) \int_w^y p(x, \omega_2^0)\, d\mu(x) > 0. \tag{18}$$

Eq. (17) implies $\int_w^y p(x, \omega_1^0)\, d\mu(x) > \int_{\bar{x}_1(w)}^{x_1(y)} p(x, \omega_1^0)\, d\mu(x)$, but this contradicts (18). Therefore, $D_2(\omega_2^0; x_1(y), y, y) \leqq 0$.

(c.3) For every $x_2 \in [v, y]$, $D_1(\omega_1^0; x_1, x_2, y) = 0$ has a solution. By continuity, then, there exists an $x_1^{**}(y), x_2^{**}(y)$ such that $D_1(\omega_1^0; x_1^{**}(y), x_2^{**}(y), y) = D_2(\omega_2^0; x_1^{**}(y), x_2^{**}(y), y) = 0$.

(a), (b), and (c) of (5) show that there exists a 3-tuple $(x_1^3(z), x_2^3(z), x_3^3(z))$ which satisfies $D_i(\omega_i^0; x_1^3, x_2^3, x_3^3) = 0$, $i = 1, 2, 3$.

The steps for $x_1 \leqq x_2 \leqq x_3 \leqq x_4 \leqq x_5 = \cdots = z$ utilize the same principles as those employed above. The general pattern should now be clear to the reader.

The next step would consider the four functions $E_i(\omega; x_1, x_2, x_3, x_4) = B_i^\varphi(\omega)$, $i = 1, \cdots, 4$ where $\varphi \sim (x_1, x_2, x_3, x_4, z, z, \cdots, z)$. It is necessary to show that $E_i(\omega_i^0) = 0$, $i = 1, 2, 3, 4$, have a solution in x_1, x_2, x_3, and x_4 This entails repeating the entire preceding argument for the case of one, two, and three functions in each case using a suitable comparison monotone procedure. We sketch the argument. Setting $x_4 = z$ we obtain by (5) that there exists a tuple $(x_1(z), x_2(z), x_3(z), z)$ for which $E_i(\omega_i^0; x_1(z), x_2(z), x_3(z), z) = 0$ for $i = 1, 2, 3$. Corollary 3 may be applied by using the second procedure $\varphi^0 \sim (x_1^0, x_2^0, \cdots, x_n^0)$ to show that $E_4(\omega_4^0; x_1(z), x_2(z), x_3(z), z) \leqq 0$. Next put $x_3 = x_4 = t < z$ and again by (5) we obtain a tuple $(x_1(t), x_2(t), t, t)$ for which $E_i(\omega_i^0; x_1(t), x_2(t), t, t) = 0$ for $i = 1, 2, 3$. According to Lemma 7, $E_4(\omega_4^0; x_1(t), x_2(t), t, t) \geqq 0$. Given y, $t < y < z$, it would be enough to construct a solution to $E_i(\omega_i^0; x_1, x_2, x_3, y) = 0$, $i = 1, 2, 3$, for then by continuity there would exist a solution to $E_i(\omega_i^0) = 0$, $i = 1, 2, 3, 4$. The analysis of $E_i(\omega_i^0; x_1, x_2, x_3, y)$, $i = 1, 2, 3$, is similar to the arguments of (5) this time using the comparison procedure

$$\varphi \sim (x_1(t), x_2(t), t, t, z, z, \cdots, z)$$

as $\varphi \sim (\bar{x}_1(w), w, w, z, \cdots, z)$ was used in (5). For the final step we repeat this sequence of arguments $n - 1$ times. This completes the proof of Theorem 1.

Corollary 4. *The unique monotone unbiased procedure defined by Theorem 1 is non-degenerate.*

Remark. Since the density $p(x, \omega)$ is assumed to have a strict monotone likelihood ratio, the set $\sigma_\omega = \{x \mid p(x, \omega) > 0\}$ is independent of ω [4]. The concept of an interval (x_i, x_{i+1}) being degenerate should therefore be understood as taken with respect to $d\mu(x)$.

Proof. Suppose the unique unbiased procedure $x^0 = (x_0, x_1, x_2, \cdots, x_{n-1}, x_n)$ where $x_0 = -\infty$ and $x_n = +\infty$ possesses a degenerate interval. We shall prove that this assumption leads to an absurdity. First, observe that (x_0, x_1) must be non-degenerate. Otherwise, let j_0 be such that (x_{j_0}, x_{j_0+1}) is the first non-degenerate interval and $j_0 \geqq 1$. By condition (b), $B_{j_0-1}^{x^0}(\omega_{j_0-1}^0) > 0$, which contradicts the definition of x^0. Now let i_0 be the earliest interval where (x_{i_0}, x_{i_0+1}) is degenerate. Therefore by what has been established $i_0 \geqq 1$ and also $i_0 < n - 1$ for in the contrary case $B_{n-1}^{x^0}(\omega_{n-1}^0)$ would be negative, Let k_0 denote the smallest index larger than i_0 for which (x_{k_0}, x_{k_0+1}) is non-degenerate. A value of k must exist, for otherwise $B_{i_0}^{x^0}(\omega_{i_0}^0) < 0$.

The strict monotone likelihood ratio possessed by $p(x, \omega)$ implies that

$$(*) \qquad \int_{x_j}^{x_{j+1}} p(\xi, \omega_i^0)\, d\mu(\xi) \int_{x_r}^{x_{r+1}} p(\xi, \omega_k^0)\, d\mu(\xi) \geqq \int_{x_j}^{x_{j+1}} p(\xi, \omega_k^0)\, d\mu(\xi) \int_{x_r}^{x_{r+1}} p(\xi, \omega_i^0)\, d\mu(\xi)$$

for every $j < i_0$ and $r \geqq k_0$ with strict inequality valid for $j = i_0 - 1$ and $r = k_0$. Equation (*) in conjunction with conditions (b) and (c) and $B_{i_0}^{x^0}(\omega_{i_0}^0) = 0$ readily leads to the result

$$B_{k_0}^{x^0}(\omega_{k_0}^0) > 0,$$

which is impossible. This completes the proof.

In any special case this construction is considerably more facile than the general proof shows. We carry this out for the special case whose loss function is (II) of the preceding section. For any prescribed $x_{i-1} < x_i$ a value $x_{i+1} (x_{i+1} > x_i)$ is determined recursively, whenever possible, by

$$(19) \qquad \int_{x_{i-1}}^{x_i} p(\xi, \omega_i^0)\, d\mu(\xi) = \int_{x_i}^{x_{i+1}} p(\xi, \omega_i^0)\, d\mu(\xi)$$

for $i = 1, 2, \cdots, n - 1$ where $x_0 = -\infty$. For x_1 sufficiently near $-\infty$, it is possible to solve (19) for each x_i such that $x_i > x_{i-1}$ and each is near $-\infty$. Allowing x_1 to increase, we observe that each x_i increases; and ultimately for $x_1 < \infty$, x_n reaches ∞. Let x_i^* be the solution of (19) where $x_n^* = +\infty$. The procedure $\varphi^* \sim (x_1^*, x_2^*, \cdots, x_{n-1}^*)$ is the unique monotone unbiased procedure for the case where

$$L_i(\omega) = \begin{cases} c & \omega \not\varepsilon S_i \\ 0 & \omega \varepsilon S_i \end{cases}.$$

REFERENCES

[1] SAMUEL KARLIN, "Decision theory of Pólya type distributions. Case of two actions I," *Proceedings of the Third Berkeley Symposium on Probability and Statistics*, Vol. 1, University of California Press, 1955, pp. 115–128.

[2] SAMUEL KARLIN, "Pólya type II" "Pólya type distributions, II," *Ann. Math. Stat.*, vol. 28 (1957), pp. 281–308

[3] SAMUEL KARLIN, "Pólya type III. Admissibility for multi-action problems," *Ann. Math. Stat.*, vol. 28 (1957), pp. 839–860.

[4] SAMUEL KARLIN and HERMAN RUBIN, "The theory of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, vol. 27 (1956), pp. 272–300.

[5] DAVID BLACKWELL and M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, 1954.

[6] L. SAVAGE, *The Foundations of Statistics*, John Wiley and Sons, 1954.

[7] E. L. LEHMANN, "A general concept of unbiasedness," *Ann. Math. Stat.*, vol. 22 (1951), pp. 587–592.

[8] E. L. LEHMANN, "A theory of some multiple decision problems, I," *Ann. Math. Stat* , vol. 28 (1957), pp. 1–25.

THE USE OF GROUP DIVISIBLE DESIGNS FOR CONFOUNDED ASYMMETRICAL FACTORIAL ARRANGEMENTS[1]

By Marvin Zelen

National Bureau of Standards

1. Introduction and summary. A factorial experiment involving m factors such that the ith factor has m_i levels is termed an asymmetrical factorial design. If the number of levels is equal to one another the experiment is termed a symmetric factorial experiment. When the block size of the experiment permits only a sub-set of the factorial combinations to be assigned to the experimental units within a block, resort is made to the theory of confounding. With respect to symmetric factorial designs, the theory of confounding has been highly developed by Bose [1], Bose and Kishen [4], and Fisher [11], [12]. An excellent summary of the results of this research appears in Kempthorne [13]. However, these researches are closely related to Galois field theory resulting in (i) only symmetric factorial designs being incorporated into the current theory of confounding; (ii) the common level must be a prime (or power of a prime) number; and (iii) the block size must be a multiple of this prime number.

The theory of confounding for asymmetric designs has not been developed to any great degree. Examples of asymmetric designs can be found in Yates [19], Cochran and Cox [9], Li [15], and Kempthorne [13]. Nair and Rao [16] have given the statistical analysis of a class of asymmetrical two-factor designs in considerable detail.

Kramer and Bradley [14] discuss the application of group divisible designs to asymmetrical factorial experiments, however their paper is mainly confined to the two-factor case and its intra-block analysis.[2] It is the purpose of this paper, which was done independently of their work, to outline the general theory for using the group divisible incomplete block designs for asymmetrical factorial experiments.

The use of incomplete block designs for asymmetric factorial experiments results in (i) no restriction that the levels must be a prime (or power of a prime) number, (ii) no restriction with respect to the dependence of the block size on the type of level, and (iii) unlike the previous referenced works on asymmetric factorial designs, the resulting analysis is simple, does not increase in difficulty with an increasing number of factors, and "automatically adjusts" for the effects of partial confounding.

Received January 10, 1957; revised June 18, 1957; revised November 1, 1957.

[1] This paper is an extension of results presented at the Annual Meeting of the American Statistical Society, September, 1954 (cf. [22]).

[2] *Note added in proof:* The Editor has pointed out that the paper by K. R. Nair, "A note on group divisible incomplete block designs", *Calcutta Statistical Association Bulletin*, Vol. 5, No. 17, (1953), pp. 30–35, together with Nair and Rao [16] essentially contains the results for the intra-block analysis of the two-factor asymmetrical designs.

Section 2 states three useful lemmas, Section 3 contains the main results of this paper, and Section 4 outlines the recovery of inter-block information.

2. Some useful lemmas.

We state here three lemmas which will be referred to in later sections. Since the proofs are trivial they are omitted.

Let $X' = (X_1, X_2, \cdots, X_n)$ have a multivariate normal distribution such that

$$E(X') = m' = (m_1, m_2, \cdots, m_n),$$

$$E[(X - m)(X - m)'] = M\sigma^2.$$

LEMMA 2.1. *The expected value of the quadratic form $X'AX$ is*

$$E(X'AX) = m'Am + \sigma^2 \text{ trace } (AM).$$

LEMMA 2.2. *If $M^2 = \lambda M$ (λ a scalar), then the quadratic form*

$$\frac{(X - m)'(X - m)}{\lambda}$$

follows a $\sigma^2\chi^2$ distribution with r degrees of freedom where $r \leqq n$ is the rank of M.

LEMMA 2.3. *Define the direct-product of two square matrices A and B of dimensions m and n respectively by*

$$(A^*B) = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mm}B \end{bmatrix}.$$

If $A^2 = \alpha A$ and $B^2 = \beta B$ (α and β are scalars), then $(A * B)^2 = \alpha\beta(A * B)$. In general, given p matrices $A, B, C, \cdots$ such that $A^2 = \alpha A$, $B^2 = \beta A$, $C^2 = \gamma C$, $\cdots$ we have $(A * B * C * \cdots)^2 = (\alpha\beta\gamma \cdots)(A * B * C * \cdots)$.

3. Analysis of group divisible designs used as asymmetrical factorials.

3.1. Estimation. The group divisible designs are partially balanced incomplete block designs with two associate classes. These were first discussed extensively by Bose and Connor [3] and Bose and Shimamato [5]. A large catalogue of such experiment plans giving full details of the analysis can be found in Bose, Clatworthy, and Shrikhande [2]. Designs with block size $k = 2$ have been enumerated by Clatworthy [7]. Bose, Shrikhande, and Battacharya [6], and Clatworthy [8] give methods for constructing group divisible designs.

Briefly group divisible designs can be characterized by having b blocks with

k experimental units such that each of the $v = mn$ treatments is replicated r times. The $v = mn$ treatments can be divided into m groups of n treatments each, where any two treatments in the same group are 1st associates and two treatments in different groups are 2nd associates. With respect to any treatment, there will be $(n - 1)$ 1st associates and $n(m - 1)$ 2nd associates.

Consider a factorial experiment with $(g + h)$ factors $A_1, A_2, \cdots, A_g, B_1, \cdots, B_h$ such that the number of levels associated with A_s is m_s for $s = 1, \cdots, g$ and the number of levels associated with B_r is n_r $r = 1, 2, \cdots, h$. Furthermore, let these levels be such that $m = \prod_{s=1}^{g} m_s$ and $n = \prod_{r=1}^{h} n_r$. Then one can use the group divisible designs for a $\prod_{s=1}^{g} m_s \times \prod_{r=1}^{h} n_r$ factorial design by arranging the $v = mn$ treatments in an $n \times m$ array and assigning the m factorial combinations among the A factors to the columns (groups) and the n factorial combinations among the B factors to the rows.

Let the measurement of the uth treatment combination $(u = 1, 2, \cdots, v)$ measured in the zth block be denoted by y_{uz} and let the underlying mathematical model be

$$y_{uz} = m + t_u + b_z + \epsilon_{uz}, \tag{3.1}$$

where m is a constant common to all measurements, t_u is the effect of the uth treatment combination, b_z is the constant associated with the zth block

$$(z = 1, 2, \cdots, b),$$

and $\{\epsilon_{uz}\}$ is a sequence of uncorrelated random variables having a zero mean and (unknown) variance σ^2. For making all tests of significance, we shall further assume that the $\{\epsilon_{uz}\}$ follow a normal distribution.

Due to the factorial nature of the experiment, a treatment combination t_z can be written as

$$\begin{aligned} t_z = {} & \sum_{s=1}^{g} (a_s)_{i_s} + \sum_{q=1}^{h} (b_q)_{j_q} \\ & + \sum_{t=2}^{g} \sum_{s=1}^{t} (a_{st})_{i_{st}} + \sum_{r=2}^{h} \sum_{q=1}^{r} (b_{qr})_{j_{qr}} \\ & + \sum_{q=1}^{h} \sum_{s=1}^{g} (a_s b_q)_{i_s j_q} + \cdots + (a_{12\ldots g} b_{12\ldots h})_{i_{12\cdots g} j_{12\cdots h}}. \end{aligned} \tag{3.2}$$

The $(a_s)_{i_s}$ are constants associated with the main effect of A_s at level i_s; the $(a_{st})_{i_{st}}$ are constants associated with the two factor interaction between A_s and A_t at levels i_s and i_t, etc. Similar interpretations hold for the constants associated with the main effects and interactions of the B factors, and also for the constants associated with the interactions composed of both A and B factors. It is well known that these parameters are not all linearly independent and

satisfy the following relations:

$$(3.3)\quad \begin{cases} \sum_{i_s=1}^{m_s} (a_s)_{i_s} = 0, & s = 1, 2, \cdots, g, \\ \sum_{j_q=1}^{n_q} (b_q)_{j_q} = 0, & q = 1, 2, \cdots, h, \\ \sum_{i_\alpha=1}^{m_\alpha} (a_{st})_{i_{st}} = 0, & \alpha = s, t;\ s < t = 1, 2, \cdots, g, \\ \sum_{j_\beta=1}^{n_\beta} (b_{qr})_{j_{qr}} = 0, & \beta = q, r;\ q < r = 1, 2, \cdots, h, \\ \vdots & \vdots \\ \sum_{i_\alpha=1}^{m_\alpha} (a_{12\cdots g} b_{12\cdots h})_{i_{12\cdots g} j_{12\cdots h}} & \\ = \sum_{j_\beta=1}^{n_\beta} (a_{12\cdots g} b_{12\cdots h})_{i_{12\cdots g} j_{12\cdots h}} = 0, & \alpha = 1, 2, \cdots g;\ \beta = 1, 2, \cdots h. \end{cases}$$

If the adjusted treatment total for the uth treatment is defined by

$$Q_u = (u\text{th treatment total}) - \begin{pmatrix}\text{sum of the block averages in} \\ \text{which the } u\text{th treatment occurs}\end{pmatrix},$$

then the treatment estimates can conveniently be written as

$$(3.4)\qquad \hat{t}_u = \frac{1}{r(k-1)} [kQ_u + c_1 S_1(Q_u) + c_2 S_2(Q_u)].$$

Here $S_1(Q_u)$ and $S_2(Q_u)$ are the sum of the adjusted treatment totals for the 1st and 2nd associates with respect to treatment u, and c_1, c_2 are constants calculated from the design parameters. (All catalogues of group divisible designs [2], [5], [7], [8], give numerical values of c_1 and c_2).

Since these estimates satisfy the restraint $\sum_{u=1}^{v} \hat{t}_u = 0$, the variance of a treatment estimate can be written as

$$(3.5)\qquad \text{Var } \hat{t}_u = \left[\frac{vk - [k + (n-1)c_1 + n(m-1)c_2]}{r(k-1)v}\right] \sigma^2,$$

and the covariance between treatments which are (say) sth associates ($s = 1, 2,$) is

$$(3.6)\qquad \text{Cov } (\hat{t}_i, \hat{t}_j) = \left[\frac{vc_s - [k + (n-1)c_1 + n(m-1)c_2]}{r(k-1)v}\right] \sigma^2$$

for $s = 1, 2$.

Let (say) $A_1, A_2, \cdots, A_p$ ($p \leqq g$) and $B_1, B_2, \cdots, B_q$ ($q \leqq h$) be a selection of the A and B factors and let them be associated with the particular

k experimental units such that each of the $v = mn$ treatments is replicated r times. The $v = mn$ treatments can be divided into m groups of n treatments each, where any two treatments in the same group are 1st associates and two treatments in different groups are 2nd associates. With respect to any treatment, there will be $(n - 1)$ 1st associates and $n(m - 1)$ 2nd associates.

Consider a factorial experiment with $(g + h)$ factors $A_1, A_2, \cdots, A_g, B_1, \cdots, B_h$ such that the number of levels associated with A_s is m_s for $s = 1, \cdots, g$ and the number of levels associated with B_r is n_r $r = 1, 2, \cdots, h$. Furthermore, let these levels be such that $m = \prod_{s=1}^{g} m_s$ and $n = \prod_{r=1}^{h} n_r$. Then one can use the group divisible designs for a $\prod_{s=1}^{g} m_s \times \prod_{r=1}^{h} n_r$ factorial design by arranging the $v = mn$ treatments in an $n \times m$ array and assigning the m factorial combinations among the A factors to the columns (groups) and the n factorial combinations among the B factors to the rows.

Let the measurement of the uth treatment combination $(u = 1, 2, \cdots, v)$ measured in the zth block be denoted by y_{uz} and let the underlying mathematical model be

$$y_{uz} = m + t_u + b_z + \epsilon_{uz}, \tag{3.1}$$

where m is a constant common to all measurements, t_u is the effect of the uth treatment combination, b_z is the constant associated with the zth block

$$(z = 1, 2, \cdots, b),$$

and $\{\epsilon_{uz}\}$ is a sequence of uncorrelated random variables having a zero mean and (unknown) variance σ^2. For making all tests of significance, we shall further assume that the $\{\epsilon_{uz}\}$ follow a normal distribution.

Due to the factorial nature of the experiment, a treatment combination t_z can be written as

$$\begin{aligned} t_z = {} & \sum_{s=1}^{g} (a_s)_{i_s} + \sum_{q=1}^{h} (b_q)_{j_q} \\ & + \sum_{t=2}^{g} \sum_{s=1}^{t} (a_{st})_{i_{st}} + \sum_{r=2}^{h} \sum_{q=1}^{r} (b_{qr})_{j_{qr}} \\ & + \sum_{q=1}^{h} \sum_{s=1}^{g} (a_s b_q)_{i_s j_q} + \cdots + (a_{12\cdots g} b_{12\cdots h})_{i_{12\cdots g} j_{12\cdots h}}. \end{aligned} \tag{3.2}$$

The $(a_s)_{i_s}$ are constants associated with the main effect of A_s at level i_s; the $(a_{st})_{i_{st}}$ are constants associated with the two factor interaction between A_s and A_t at levels i_s and i_t, etc. Similar interpretations hold for the constants associated with the main effects and interactions of the B factors, and also for the constants associated with the interactions composed of both A and B factors. It is well known that these parameters are not all linearly independent and

satisfy the following relations:

$$
(3.3)\quad
\begin{cases}
\sum_{i_s=1}^{m_s} (a_s)_{i_s} = 0, & s = 1, 2, \cdots, g, \\
\sum_{j_q=1}^{n_q} (b_q)_{j_q} = 0, & q = 1, 2, \cdots, h, \\
\sum_{i_\alpha=1}^{m_\alpha} (a_{st})_{i_s i_t} = 0, & \alpha = s, t; \quad s < t = 1, 2, \cdots, g, \\
\sum_{j_\beta=1}^{n_\beta} (b_{qr})_{j_q j_r} = 0, & \beta = q, r; \quad q < r = 1, 2, \cdots, h, \\
\quad \vdots & \quad \vdots \\
\sum_{i_\alpha=1}^{m_\alpha} (a_{12\cdots g} b_{12\cdots h})_{i_1 i_2 \cdots i_g j_1 j_2 \cdots j_h} & \\
= \sum_{j_\beta=1}^{n_\beta} (a_{12\cdots g} b_{12\cdots h})_{i_1 i_2 \cdots i_g j_1 j_2 \cdots j_h} = 0, & \alpha = 1, 2, \cdots g; \quad \beta = 1, 2, \cdots h.
\end{cases}
$$

If the adjusted treatment total for the uth treatment is defined by

$$Q_u = (u\text{th treatment total}) - \begin{pmatrix}\text{sum of the block averages in} \\ \text{which the } u\text{th treatment occurs}\end{pmatrix},$$

then the treatment estimates can conveniently be written as

$$(3.4)\qquad \hat{t}_u = \frac{1}{r(k-1)} [kQ_u + c_1 S_1(Q_u) + c_2 S_2(Q_u)].$$

Here $S_1(Q_u)$ and $S_2(Q_u)$ are the sum of the adjusted treatment totals for the 1st and 2nd associates with respect to treatment u, and c_1, c_2 are constants calculated from the design parameters. (All catalogues of group divisible designs [2], [5], [7], [8], give numerical values of c_1 and c_2).

Since these estimates satisfy the restraint $\sum_{u=1}^{v} \hat{t}_u = 0$, the variance of a treatment estimate can be written as

$$(3.5)\qquad \operatorname{Var} \hat{t}_u = \left[\frac{vk - [k + (n-1)c_1 + n(m-1)c_2]}{r(k-1)v}\right] \sigma^2,$$

and the covariance between treatments which are (say) sth associates ($s = 1, 2,$) is

$$(3.6)\qquad \operatorname{Cov}(\hat{t}_i, \hat{t}_j) = \left[\frac{vc_s - [k + (n-1)c_1 + n(m-1)c_2]}{r(k-1)v}\right] \sigma^2$$

for $s = 1, 2$.

Let (say) $A_1, A_2, \cdots, A_p$ ($p \leqq g$) and $B_1, B_2, \cdots, B_q$ ($q \leqq h$) be a selection of the A and B factors and let them be associated with the particular

levels $i = (i_1, i_2, \cdots, i_p)$ and $j = (j_1, j_2, \cdots, j_q)$ respectively. We define an S-function associated with these particular factors and levels by

$$S[1, 2, \cdots, p;\quad 1, 2, \cdots, q \mid i, j] = \frac{\prod_{i=1}^{p} m_i \prod_{j=1}^{q} n_j}{v} \sum_u{}' \hat{t}_u, \tag{3.7}$$

where the summation $\sum_u'$ refers to the sum over all treatment estimates which have the same levels $i_1, i_2, \cdots, i_p; j_1, j_2, \cdots, j_q$ with respect to the factors $A_1, A_2, \cdots, A_p; B_1, B_2, \cdots, B_q$. If an S-function contains no A factors, we shall denote this by $S[0; 1, 2, \cdots, q \mid j]$ with a similar notation for the absence of B factors. (Note that these S-functions are simply the cell averages in any $(p + q)$ way table associated with these factors). Then the expected value of (3.7) is

$$\begin{aligned} E\{S[1, 2, \cdots, p;\quad 1, 2, \cdots, q \mid i, j]\} &= \sum_{s=1}^{p} (a_s)_{i_s} + \sum_{r=1}^{q} (b_r)_{j_r} \\ &+ \sum_{t=2}^{p} \sum_{s=1}^{t} (a_{st})_{i_{st}} + \sum_{s=2}^{q} \sum_{r=1}^{s} (b_{rs})_{j_{rs}} + \cdots + (a_{12\cdots p} b_{12\cdots q})_{i_{12\cdots p} j_{12\cdots q}}, \end{aligned} \tag{3.8}$$

where the summations refer to the particular factors A_i $(i = 1, 2, \cdots, p)$, B_j $(j = 1, 2, \cdots, q)$ and the levels $i_{st\cdots} j_{rs\cdots}$ refer only to $i = (i_1, i_2, \cdots, i_p)$ and $j = (j_1, j_2, \cdots, j_q)$. There will be only $(v - 1)$ linearly independent treatment estimates and since the relations (3.3) imply that there exist $(v - 1)$ linearly independent factorial constants, the condition of unbiasedness is sufficient to insure unique estimates of the factorial constants. Therefore the estimates of the main effects and interaction parameters are given by

$$\left\{\begin{aligned} &(\hat{a}_s)_{i_s} = S[s; 0 \mid i_s], \\ &(\hat{b}_q)_{j_q} = S[0; q \mid j_q], \\ &(\hat{a}_{st})_{i_{st}} = S[s, t; 0 \mid i_s, i_t] - \{S[s; 0 \mid i_s] + S[t; 0 \mid i_t]\}, \\ &(\hat{b}_{qu})_{j_{qu}} = S[0; q, u \mid j_q, j_u] - \{S[0; q \mid j_q] + S[0; u \mid j_u]\}, \\ &(a_s\hat{b}_q)_{i_s j_q} = S[s; q \mid i_s, j_q] - \{S[s; 0 \mid i_s] + S[0; q \mid j_q]\}, \\ &\quad \vdots \qquad\qquad\qquad\qquad \vdots \\ &(a_{12\cdots g}\hat{b}_{12\cdots h})_{i_{12\cdots g} j_{12\cdots h}} \\ &\quad = S[1, 2, \cdots, g; 1, 2, \cdots, h \mid i_1, \cdots, i_g, j_1, \cdots, j_h] \\ &\quad - \{S[1, 2, \cdots, g - 1; 1, 2, \cdots, h \mid i_1, \cdots, i_{g-1}, j_1, \cdots, j_h] + \cdots\}. \\ &\quad + \cdots + (-1)^{g+h-1}\{S[1; 0 \mid i_1] + \cdots\}. \end{aligned}\right. \tag{3.9}$$

The estimate for a $(p + q)$th interaction involving the factors (say) $\{A_s\}$, $\{B_r\}$ associated with the respective levels i_s, j_r $(s = 1, 2, \cdots, p; r = 1, 2, \cdots, q)$ can conveniently be written as

$$(a_{12\cdots p}\hat{b}_{12\cdots q})_{i_{12\cdots p} j_{12\cdots q}} = (-1)^{p+q} \sum_{w=1}^{p+q} (-1)^w \{w\}, \tag{3.10}$$

where $\{w\}$ denotes the sum of all S-functions involving exactly $w \leqq p + q$ factors from the above set.

3.2. Variances, covariances, and tests of significance. In this section we shall obtain the variances and covariances of the main effects and interaction terms. It will be shown that these can be written as direct products of matrices and this leads directly to the appropriate sums of squares for the analysis of variance. Four lemmas pertaining to the S-functions are derived and are used for proving three basic theorems pertaining to the analysis.

LEMMA 3.1. *The variance of* $S = S[1, 2, \cdots, p; 1, 2, \cdots, q \mid i, j]$ *is*

$$\text{(3.11)} \qquad \operatorname{Var} S = \frac{\sigma^2}{r(k-1)v}\,[(MN-1)(k-c_1) + n(M-1)(c_1-c_2)],$$

where $M = \prod_{s=1}^{p} m_s$, $N = \prod_{r=1}^{q} n_r$.

PROOF. The number of treatments summed in S is $v/MN = mn/MN$ which can be regarded as m/M groups of n/N treatments each, such that treatments within the same group are first associates and treatments in different groups are second associates. Then there are $\binom{mn/MN}{2}$ different pairs of treatments among the mn/MN treatments in S, of which

$$\frac{m}{M}\binom{n/N}{2} = \frac{v(n-N)}{2N^2M}$$

are 1st associates and

$$\frac{1}{2}\left(\frac{m}{M}-1\right)\left(\frac{n}{N}\right)\frac{v}{NM}$$

are 2nd associates. Therefore the variance of S is

$$\text{(3.12)} \quad \left\{\begin{aligned} \operatorname{Var} S = \frac{\sigma^2M^2N^2}{v^2}\bigg\{&\frac{v}{MN}\left[\frac{vk-[k+c_1(n-1)+c_2n(m-1)]}{r(k-1)v}\right] \\ &+\frac{2v(n-N)}{2N^2M}\left[\frac{c_1v-[k+c_1(n-1)+c_2n(m-1)]}{r(k-1)v}\right] \\ &+\frac{2v(m-M)n}{2(MN)^2}\left[\frac{c_2v-[k+c_1(n-1)+c_2n(m-1)]}{r(k-1)v}\right]\bigg\}, \end{aligned}\right.$$

which on simplifying gives the desired result.

LEMMA 3.2. *Let* $S = S[1, 2, \cdots, p; 1, 2, \cdots, q \mid i, j]$ *and*

$$S' = S[1', 2', \cdots, p'; 1', 2', \cdots, q' \mid i', j']$$

be two S-functions having a A factors and b B factors in common, such that for a_1 *and* b_1 *of these factors, the levels are identical and for* a_2 *and* b_2 *of these (common) factors, the levels are different* ($a = a_1 + a_2$, $b = b_1 + b_2$). *Then*

$$\text{(3.13)} \qquad \operatorname{Cov}(S, S') = \frac{\sigma^2}{r(k-1)v}\left\{\begin{aligned}&(M_1N_1-1)(k-c_1)\\ &+n(M_1-1)(c_1-c_2)\end{aligned}\right\},$$

where $M_1 = \prod_{i=1}^{a_1} m_i$ *(product of the levels of the* a_1 *factors having common levels) and* $N_1 = \prod_{j=1}^{b_1} n_j$ *(product of the levels of the* b_1 *factors having common levels).*

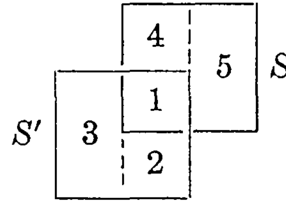
PROOF. The number of treatments summed in S and S' are v/MN and $v/M'N'$ respectively. These treatments can be regarded as consisting of two rectangular treatment arrays of dimensions $(n/N) \times (m/M)$ and

$$(n/N') \times (m/M')$$

respectively. The two arrays will overlap if they have common treatments and the number of such common treatments is

$$\frac{vM_1N_1}{(MN)(M'N')} = \left(\frac{mM_1}{MM'}\right)\left(\frac{nN_1}{NN'}\right).$$

It is convenient to depict the intersection of the rectangular arrays by the five regions as shown below,



where region (1) is an array representing the common treatments having (nN_1/NN') rows and (mM_1/MM') columns. If $\sum (i)$ represent the sum of the treatments in the ith region ($i = 1, 2, 3, 4, 5$), then

$$\text{(3.14)} \qquad \begin{cases} S = \sum (1) + \sum (4) + \sum (5), \\ S' = \sum (1) + \sum (2) + \sum (3). \end{cases}$$

Hence, in order to find the covariance between S and S', it is necessary to find the number of pairs of 1st and 2nd associates formed from the multiplication of S and S'. These will give pairs formed from $\sum (1)^2$, $\sum (1) \sum (2)$, $\sum (1) \sum (3)$, $\sum (1) \sum (4)$, $\sum (2) \sum (4)$, $\sum (3) \sum (4)$, $\sum (1) \sum (5)$, $\sum (2) \sum (5)$, and $\sum (3) \sum (5)$.

Define

$$\text{(3.15)} \qquad \begin{cases} m_1 = \dfrac{mM_1}{MM'}, & n_1 = \dfrac{nN_1}{NN'}, \\[2ex] m_2 = \dfrac{m}{MM'}(M - M_1), & n_2 = \dfrac{n}{NN'}(N - N_1), \\[2ex] m_3 = \dfrac{m}{MM'}(M' - M_1), & n_3 = \dfrac{n}{NN'}(N' - N_1). \end{cases}$$

Then the dimensions of the five regions are:

$$\text{region (1):} \quad n_1 \times m_1,$$

$$\text{region (2):} \quad n_2 \times m_2,$$

region (3): $(n_1 + n_2) \times m_2$,

region (4): $n_3 \times m_1$,

region (5): $(n_1 + n_3) \times m_3$.

Since the treatments in the same row are 1st associates of each other and treatments in different rows 2nd associates, it is an easy matter to count the number of 1st and 2nd associates arising from pairs formed from $\sum (i) \sum (j)$. Performing the necessary algebra, we find that the total number of 1st associate pairs is

$$\frac{vM_1(n - N_1)}{(NN')(MM')}$$

and the total number of 2nd associate pairs is

$$\frac{vn(m - M_1)}{(NN')(MM')}.$$

Therefore,

$$(3.16) \quad \begin{cases} \operatorname{Cov}(S, S') = \dfrac{\sigma^2(MM')(NN')}{v^2}\left\{\dfrac{vM_1N_1}{(MM'NN')} \quad [\operatorname{Var} t] \right. \\ \qquad + \dfrac{vM_1(n - N_1)}{(NN'MM')} \quad \operatorname{Cov}\ (\text{1st associates}) \\ \qquad \left. + \dfrac{vn(m - M_1)}{(NN'MM')} \quad \operatorname{Cov}\ (\text{2nd associates})\right\}. \end{cases}$$

On simplifying we get the desired result.

LEMMA 3.3. *Let* $(\hat{ab}) = (a_{12\cdots p}\hat{b}_{12\cdots q})_{i_{12}\cdots_p j_{12}\cdots_q}$ *be the estimate of the* $(p + q)$th *factor interaction associated with the factors* $\{A_s\}(s = 1, 2, \cdots, p)$ *and*

$$\{B_r\}(r = 1, 2, \cdots, q).$$

Let $S' = S[1; 2', \cdots, p'; 1', 2', \cdots, q' \mid i', j']$ *be an S-function which is not associated with all factors (regardless of level) of* $(\hat{ab})$. *Then*

$$(3.17) \qquad \operatorname{Cov}[(\hat{ab}), S'] = 0.$$

PROOF. Let α be the number of common A factors between (ab) and S', and α_1 and α_2 ($\alpha = \alpha_1 + \alpha_2$) be the number of these common factors having the same levels and different levels, respectively. Define β, β_1, and β_2 in the same manner with respect to the B factors. Since the interaction $(\hat{ab})$ can be written in the form

$$(\hat{ab}) = (-1)^{p+q} \sum_{w=1}^{p+q} (-1)^w \{w\},$$

consider a fixed $\{w\}$ and a particular S-function in $\{w\}$ having the characteristics a, a_1, a_2, b, b_1, b_2 as defined in Lemma 3.2.

Define

$$(3.18)\quad \begin{cases} C(0, 0) = -1, \\ C(a_1) = \sum \cdots \sum (m_{s_1} m_{s_2} \cdots m_{s_{a_1}} - 1), & a_1 \leqq \alpha_1 , \\ C(b_1) = \sum \cdots \sum (n_{r_1} n_{r_2} \cdots n_{r_b} - 1), & b_1 \leqq \beta_1 , \\ C(a_1 , b_1) = \sum \cdots \sum (m_{s_1} m_{s_2} \cdots m_{s_{a_1}} n_{r_1} n_{r_2} \cdots n_{r_{b_1}} - 1), \\ \qquad\qquad a_1 \leqq \alpha_1 , \quad b_1 \leqq \beta_1 , \end{cases}$$

where the summations are only over combinations of A and B factors taken a_1 and b_1 at a time respectively, such that these factors are those in which $(\widehat{ab})$ and S have in common at the same level.

Then the covariance between S' and $\{w\}$ can be written

$$(3.19)\quad \begin{cases} \operatorname{Cov}[S', \{w\}] = \dfrac{\sigma^2}{r(k-1)v} \left\{ \displaystyle\sum_{a_1+b_1=w} \binom{p+q-\alpha-\beta}{w-a_1-b_1} \right. \\ \qquad \cdot \left[(k-c_1)C(a_1, b_1) + n \dbinom{\beta_1}{b_1} (c_1 - c_2) C(a_1) \right] \\ - \displaystyle\sum_{\substack{a_1+b=w \\ b_2 \neq 0}} \binom{p+q-\alpha-\beta}{w-a_1-b} \\ \qquad \cdot \left[\dbinom{\alpha_1}{a_1} \dbinom{\beta}{b} (k-c_1) - n \dbinom{\beta}{b} (c_1 - c_2) C(a_1) \right] \\ \left. + \displaystyle\sum_{a_2=1}^{w} \binom{p+q-\alpha-\beta}{w-a_2} \binom{\alpha_2}{a_2} [(k-c_1) + n(c_1-c_2)] \right\}. \end{cases}$$

Note that the first summation is for those S-functions in $\{w\}$ for which

$$a_2 = b_2 = 0;$$

the second summation refers to $a_2 = 0$, $b_2 \neq 0$; and the third summation is when $a_2 \neq 0$. Since the covariance between S' and $(\widehat{ab})$ is

$$(3.20)\qquad \operatorname{Cov}[S', (\widehat{\ })] = (-1)^{p+q} \sum_{w=1}^{p+q} (-1)^w \operatorname{Cov}[S', \{w\}],$$

we can substitute (3.19) in (3.20) to obtain an explicit expression for (3.20). Now with respect to fixed values of a_1, a_2, b_1, and b_2 the only terms contributing to the first summation in (3.19) is when

$$w = a_1 + b_1 , \cdots , p + q + a_1 + b_1 - \alpha - \beta;$$

the value of w contributing to the second summation in (3.19) is for

$$w = a_1 + b, \cdots , p + q + a_1 + b - \alpha - \beta$$

and the contributing value of w for the last summation in (3.19) is when

$$w = a_2, \cdots, p + q + a_2 - \alpha - \beta.$$

Therefore collecting coefficients of

$$\left[(k - c_1)C(a_1, b_1) + n\binom{\beta_1}{b_1}(c_1 - c_2)C(a_1)\right]$$

in (3.20) gives

$$(3.21) \qquad (-1)^{a_1+b_1} \sum_{w=0}^{p+q-\alpha-\beta} \binom{p + q - \alpha - \beta}{w} (-1)^w = 0$$

for all a_1 and b_1. Collecting coefficients of

$$\binom{\beta}{b}\left[\binom{\alpha_1}{a_1}(k - c_1) - n(c_1 - c_2)C(a_1)\right]$$

results in

$$(-1)^{a_1+b+1} \sum_{w=0}^{p+q-\alpha-\beta} \binom{p + q - \alpha - \beta}{w} (-1)^w = 0$$

for all a_1, b_1, and b_2. Finally, with respect to the coefficient of

$$\binom{\alpha_2}{a_2}[(k - c_1) + n(c_1 - c_2)]$$

in (3.20) we have

$$(-1)^{a_1} \sum_{w=0}^{p+q-\alpha-\beta} \binom{p + q - \alpha - \beta}{w} (-1)^w = 0.$$

LEMMA 3.4. *Let* $(\hat{ab}) = (a_{12\cdots p}\hat{b}_{12\cdots q})_{i_1 i_2 \cdots i_p j_1 j_2 \cdots j_q}$ *be an estimate of the* $(p + q)$ *factor interaction associated with the factors* $\{A_i\}(i = 1, 2, \cdots, p)$ *and*

$$\{B_j\}(j = 1, 2, \cdots, q).$$

Let $S' = S[1, 2, \cdots, p; 1, 2, \cdots, q \mid i, j]$ *be an S-function associated with the same factors as* $(\hat{ab})$ *such that* α_1 *and* β_1 *of the A and B factors have common levels. Then,*

$$(3.22) \quad \mathrm{Cov}\,[(\hat{ab}), S'] = \begin{cases} (-1)^{p+q+\alpha_1+\beta_1}\theta(1, 2, \cdots, \alpha_1)\phi(1, 2, \cdots, \beta_1)\dfrac{\sigma^2}{E_b r v}, & \text{if } q \neq 0, \\ (-1)^{p+\alpha_1}\theta(1, 2, \cdots, \alpha_1)\dfrac{\sigma^2}{E_a r v}, & \text{if } q = 0, \end{cases}$$

where

$$(3.23) \quad \begin{aligned} \theta(1, 2, \cdots, \alpha_1) &= (m_1 - 1)(m_2 - 1)\cdots(m_{\alpha_1} - 1), \\ \phi(1, 2, \cdots, \beta_1) &= (n_1 - 1)(n_2 - 1)\cdots(n_{\beta_1} - 1), \end{aligned}$$

and

$$(3.24)\quad \begin{cases} E_a = \dfrac{(k-1)}{(k-c_1)+n(c_1-c_2)}, \\ E_b = \dfrac{k-1}{k-c_1}. \end{cases}$$

PROOF. If we expand $(a\hat{\ }b)$ in terms of S-functions (Eq. 3.10), we can write the covariance between S' and a fixed $\{w\}$ for $q \neq 0$ as

$$(3.25)\quad \begin{aligned} \operatorname{Cov}[S', \{w\}] = \frac{\sigma^2}{r(k-1)v} \Bigg\{ & \Bigg[(k-c_1) \sum_{a_1+b_1=w} C(a_1, b_1) \\ & + n(c_1-c_2) \sum_{a_1+b_1=w} \binom{\beta_1}{b_1} C(a_1) \Bigg] \\ & - \Bigg[\sum_{\substack{a_2+s=w \\ a_2 \neq 0}} \binom{\alpha_2}{a_2}\binom{q+\alpha_1}{s} [(k-c_1)+n(c_1-c_2)] \Bigg] \\ & - \Bigg[\sum_{\substack{a_1+b_1+b_2=w \\ b_2 \neq 0}} \binom{\alpha_1}{a_1}\binom{\beta_1}{b_1}\binom{\beta_2}{b_2} [(k-c_1)] \\ & - n(c_1-c_2) \sum_{\substack{a_1+b_1+b_2=w \\ b_2 \neq 0}} \binom{\beta_1}{b_1}\binom{\beta_2}{b_2} C(a_1) \Bigg] \Bigg\}, \end{aligned}$$

where the first bracket is when $a_2 = b_2 = 0$; the second bracket is the case $a_2 \neq 0$; and the third bracket refers to $a_2 = 0$, $b_2 \neq 0$. Substituting (3.25) in

$$(3.26)\quad \operatorname{Cov}[S', (a\hat{\ }b)] = (-1)^{p+q} \sum_{w=1}^{p+q} (-1)^w \operatorname{Cov}[S', \{w\}]$$

results in the first bracket being written as (neglecting the constant term)

$$(3.27)\quad \begin{aligned} & (-1)^{p+q+\alpha_1+\beta_1} \Bigg[\sum_{w=1}^{\alpha_1+\beta_1} (-1)^w \sum_{a_1+b_1=w} C(a_1, b_1)(k-c_1) \\ & \qquad + n(c_1-c_2) \sum_{w=1}^{\alpha_1+\beta_1} (-1)^w \sum_{a_1+b_1=w} \binom{\beta_1}{b_1} C(a_1) \Bigg] \\ & = (-1)^{p+q+\alpha_1+\beta_1} \theta(1, 2, \cdots, \alpha_1)\phi(1, 2, \cdots, \beta_1)(k-c_1) \\ & \qquad + (-1)^{p+q} n(c_1-c_2) \sum_{a_1=1}^{\alpha_1} C(a_1) \sum_{w=a_1}^{a_1+q} (-1)^w \binom{\beta_1}{w-a_1}. \end{aligned}$$

With respect to the bracket when $a_2 \neq 0$, we can write these terms after substituting in (3.26) as

$$(3.28)\quad [k - c_1 + n(c_1-c_2)] \left[\sum_{r=1}^{p+q} (-1)^r \binom{p+q}{r} - \sum_{r=1}^{\alpha_1+q} (-1)^r \binom{\alpha_1+q}{r} \right] = 0.$$

Finally for the terms where $a_2 = 0$, $b_2 \neq 0$, after substituting in (3.26), we can write

$$(3.29)\quad \begin{cases} (-1)^{p+q+1}\left[\sum_{w=1}^{\alpha_1+q}(-1)^w\binom{\alpha_1+q}{w} - \sum_{w=1}^{\alpha_1+\beta_1}(-1)^w\binom{\alpha_1+\beta_1}{w}(k-c_1)\right. \\ \left. + (-1)^{p+q}n(c_1-c_2)\left[\sum_{a_1=1}^{\alpha_1}C(a_1)\sum_{w=a_1}^{\alpha_1+q}(-1)^w\left[\binom{\beta_1+\beta_2}{w-a_1}-\binom{\beta_1}{w-a_1}\right]\right]\right]. \end{cases}$$

The first term in (3.29) is identically zero and combining the second term of the right hand side of (3.27) with the second term of (3.29) gives

$$(-1)^{p+q}n(c_1-c_2)\left[\sum_{a_1=1}^{\alpha_1}C(a_1)\sum_{r=a_1}^{\alpha_1+q}(-1)^r\binom{q}{r-a_1}\right] = 0.$$

Thus the Lemma is true for $q \neq 0$. For $q = 0$, the covariance between S' and $\{w\}$ will be

$$(3.30)\quad \begin{cases} \text{Cov}\,[S', \{w\}] = \dfrac{\sigma^2}{r(k-1)v}\Big\{[(k-c_1)+n(c_1-c_2)]C(a_1) \\ \qquad - \sum_{a_1+a_2=w}\binom{\alpha_1}{a_1}\binom{\alpha_2}{a_2}[(k-c_1)+n(c_1-c_2)]\Big\} \end{cases}$$

and following the same reasoning as for $q \neq 0$, we can prove the Lemma for $q = 0$.

THEOREM 3.1. *Let* $(\hat{ab}) = (a_{12} \cdot {}_p\hat{b}_{12} \cdot {}_q)_{i_1 2\ldots p j_1 2 \ldots q}$ *be an estimate of the*

$$(p+q)\text{th}$$

factor interaction associated with the factors $\{A_i\}$ $(i = 1, 2, \cdots, p)$ *and*

$$\{B_j\}\ (j = 1, 2, \cdots, q);$$

let $(\hat{ab})' = (a'_{12}\cdots{}_r\hat{b}'_{12\cdots s})_{i'12\ldots r j'12\ldots s}$ *be a* $(r+s)$ *factor interaction associated with the factors* $\{A'_i\}$ $(i = 1, 2, \cdots, r)$ *and* $\{B'_j\}$ $(j = 1, 2, \cdots, s)$, *such that all factors are not identical between* $(\hat{ab})$ *and* $(\hat{ab})'$. *Then the two different interactions are uncorrelated.*

PROOF. $(\hat{ab})$ can be expanded in terms of S-functions, such that no S-function contains all the factors of $(\hat{ab})'$. Hence by Lemma 3.3, the covariance between all such S-functions and $(\hat{ab})'$ are zero which proves the theorem.

THEOREM 3.2. *The variance of the* $(p+q)$ *factor interaction*

$$(\hat{ab}) = (a_{12\cdots p}\hat{b}_{12\cdots q})_{i12\ldots p j12\ldots q}$$

associated with the factors $\{A_i\}$ $(i = 1, 2, \cdots, p)$ *and* $\{B_j\}$ $(j = 1, 2, \cdots, q)$ *is*

$$(3.31)\quad \text{Var}\,(\hat{ab}) = \begin{cases} \theta(1, 2, \cdots, p)\dfrac{\sigma^2}{E_a r v}, & \text{if } q = 0, \\ \theta(1, 2, \cdots, p)\phi(1, 2, \cdots, q)\dfrac{\sigma^2}{E_b r v}, & \text{if } q \neq 0. \end{cases}$$

Proof. Using Lemma 3.3, we can show that

$$\mathrm{Var}\,(\widehat{ab}) = \mathrm{Cov}\,[(\widehat{ab}),\, S],$$

where S denotes that S-function coinciding in all factors and levels with the interaction $(\widehat{ab})$. Hence, by Lemma 3.4 the theorem is proved.

Theorem 3.3. *Let $(\widehat{ab})_{ij}$ and $(\widehat{ab})_{i'j'}$ be two estimates of a $(p + q)$ factor interaction associated with the factors $\{A_i\}$ $(i = 1, 2, \cdots, p)$ and*

$$\{B_j\}\ (j = 1, 2, \cdots, q)$$

such that for α_1 of the A factors and β_1 of the B factors, the levels are identical. Then

$$\text{(3.32)}\quad \mathrm{Cov}\,[(\widehat{ab})_{ij},\, (\widehat{ab})_{i'j'}] = \begin{cases} (-1)^{p+\alpha_1}\theta(1, 2, \cdots, \alpha_1)\, \dfrac{\sigma^2}{E_a rv}, & \text{if } q = 0, \\[2ex] (-1)^{p+q+\alpha_1+\beta_1}\theta(1, 2, \cdots, \alpha_1)\phi(1, 2, \cdots, \beta_1)\, \dfrac{\sigma^2}{E_b rv}, & \text{if } q \neq 0. \end{cases}$$

Proof. Expanding $(\widehat{ab})_{i'j'}$ in terms of S-functions, taking the covariance of $(\widehat{ab})_{ij}$ with each of the S-functions associated with $(\widehat{ab})_{i'j'}$ and using Lemma 3.3, results in

$$\mathrm{Cov}\,[(\widehat{ab})_{ij},\, (\widehat{ab})_{i'j'}] = \mathrm{Cov}\,[(\widehat{ab})_{ij},\, S'],$$

where S' is that S-function associated with the factors $\{A_i\}$ and

$$\{B_j\}\ (i = 1, 2, \cdots, p;\, j = 1, 2, \cdots, q)$$

and levels $i'_{12\cdots p}j'_{12\cdots q}$.

Hence, by Lemma 3.4 the theorem is proved.

Theorems 3.1 through 3.3 give the variances and covariances of any general interaction. Define the square matrices $M(i)$ and $N(j)$ of dimension m_i and n_j, respectively, by

$$\text{(3.33)}\quad \begin{cases} M(i) = m_i I - J, & i = 1, 2, \cdots, g, \\ N(j) = n_j I - J, & j = 1, 2, \cdots, h, \end{cases}$$

where J is a matrix of appropriate dimension having all elements unity. Then the variance-covariance matrix of the estimates of the $(p + q)$ factor interaction $(a_{12\cdots p}b_{12\cdots q})_{i_{12\cdots p}j_{12\cdots q}}$ ranging over all the $\prod_{i=1}^{p} m_i \prod_{j=1}^{q} n_j$ combinations is given by the direct matrix product

$$\text{(3.34)}\quad \frac{\sigma^2}{E_a rv}\,[M(1)*M(2)*\cdots *M(p)], \qquad \text{if } q = 0,$$

or

$$\text{(3.35)}\quad \frac{\sigma^2}{E_b rv}\,[M(1)*\cdots *M(p)*N(1)*\cdots *N(q)], \qquad \text{if } q \neq 0.$$

Therefore, since $M(i)^2 = m_i M(i)$, $N(j)^2 = n_j N(j)$, and using Lemmas (2.2) and (2.3) the sums of squares

$$(3.36)\qquad \frac{E_a rv}{\prod_{s=1}^{p} m_s} \sum_{i_1 i_2 \cdots i_p} (\hat{a}_{12\cdots p})^2_{i_1 i_2 \cdots i_p},$$

$$\frac{E_b rv}{\prod_{s=1}^{p} m_s \prod_{r=1}^{q} n_r} \sum_{i_1 i_2 \cdots i_p j_1 j_2 \cdots j_q} (a_{12 \cdots p} \hat{b}_{12 \cdots q})^2_{i_1 i_2 \cdots i_p j_1 j_2 \cdots j_q}, \qquad \text{if } q \neq 0,$$

follow $\chi^2\sigma^2$ distributions with $\prod_{s=1}^{p} (m_s - 1)$ and $\prod_{s=1}^{p} \prod_{r=1}^{q} (m_s - 1)(n_r - 1)$ degrees of freedom respectively if the null hypothesis of no interaction effect is true.

Using Lemma 2.1 these sums of squares have the expected values

$$\frac{E_a rv}{\prod_{s=1}^{p} m_s} \sum_{i_1 i_2 \cdots i_p} (a_{12 \cdots p})^2_{i_1 i_2 \cdots i_p} + \prod_{i=1}^{p} (m_i - 1)\sigma^2$$

and

$$\frac{E_b rv}{\prod_{s=1}^{p} m_s \prod_{r=1}^{q} n_r} \sum_{i_1 i_2 \cdots i_p j_1 j_2 \cdots j_q} (a_{12 \cdots p} b_{12 \cdots q})^2 + \prod_{i=1}^{p} (m_i - 1) \prod_{j=1}^{q} (n_j - 1)\sigma^2.$$

The entire intra-block analysis of variance can be summarized in Table 1 where B represents the $b \times 1$ vector of block totals, Q is the $v \times 1$ vector of adjusted treatment totals, $\hat{t}$ is the $v \times 1$ vector of treatment estimates,

$$g = \frac{(\text{grand total})^2}{bk},$$

and terms such as

$$\sum_{i_1 i_2 \cdots i_p} (\hat{a}_{12 \cdots p})^2_{i_1 i_2 \cdots i_p}$$

are written as $(\hat{a}_{12\cdots p})^2$, etc.

The computations for the analysis of variance are straightforward and actually amount to treating the $\hat{t}_u$'s as observations on a one replicate factorial experiment, where all sums of squares are multiplied by $E_a r$ or $E_b r$. It is also clear from the analysis of variance that the various interactions are estimated with one of a possible two types of efficiencies. If the interaction is composed only of A factors the efficiency is E_a, otherwise the efficiency will be E_b.

Extension to the balanced incomplete block designs. The balanced incomplete block designs can also be used for asymmetric factorial arrangements by assigning the v factorial combinations to the v treatments of the balanced incomplete block design. All results immediately follow by regarding the balanced incomplete block designs as a "degenerate" partially balanced design. Then

TABLE 1

Summary of intra-block analysis of variance

Source	Degrees of freedom	Sum of squares
Blocks (unadjusted)......	$b - 1$	$\frac{B'B}{k} - g$
Treatments (adjusted)....	$v - 1$	$\hat{t}'Q$
A_1	$(m_1 - 1)$	$\frac{v}{m_1} E_a r(\hat{a}_1)^2$
A_2	$(m_2 - 1)$	$\frac{v}{m_2} E_a r(\hat{a}_2)^2$
$\vdots$	$\vdots$	$\vdots$
B_1	$(n_1 - 1)$	$\frac{v}{n_1} E_b r(\hat{b}_1)^2$
$\vdots$	$\vdots$	$\vdots$
A_1B_1	$(m_1 - 1)(n_1 - 1)$	$\frac{v}{m_1 n_1} E_b r(\widehat{a_1 b_1})^2$
$\vdots$	$\vdots$	$\vdots$
$A_1 A_2 \cdots A_g B_1 \cdots B_h$	$\prod_{s=1}^{g} (m_s - 1) \prod_{r=1}^{h} (n_r - 1)$	$\frac{v}{\prod_1^g m_s \prod_1^h n_r} E_b r(a_{12\cdots g} \hat{b}_{12\cdots h})^2$
Error.....................	$n_e = bk - b - v + 1$	$S_e = \sum_{i,j} y_{ij}^2 - t'Q - \frac{B'B}{k}$
Total................. ..	$bk - 1$	$\sum_{i,j} y_{ij}^2 - g$

$c_1 = c_2 = k/v$ in (3.4), $E_a = E_b = E = v(k-1)/k(v-1)$, and all main effects and interactions are estimated with an efficiency factor E.

4. The recovery of inter-block information. If the block effects b_j in (3.4) can be regarded as a sequence of random variables such that

$$(4.1)\qquad \begin{cases} E(b_j) = 0, \qquad \operatorname{Var} b_j = \sigma_b^2, \\ \operatorname{Cov}(b_j, b_{j'}) = 0, \qquad\qquad \text{for } j \neq j', \\ \operatorname{Cov}(\epsilon_{ij}, b_{j'}) = 0, \end{cases}$$

it will be possible to extract additional information from the block totals. This additional analysis is sometimes termed the recovery of inter-block information or the interblock analysis. With respect to the balanced incomplete block designs, the inter-block analysis was first developed by Yates [20] and appears in the books by Cochran and Cox [9], Federer [10], and Kempthorne [13]. The inter-block analysis with respect to the partially balanced designs is discussed in a particularly simple form by Bose and Shimamoto [5] and by Bose, Clatworthy, and Shrikhande [2]. Generally it will be possible to use this inter-block information in two ways. The preceding references discuss how one may combine the inter-block information with the intra-block information in order to

obtain the most precise treatment estimates. The paper by Zelen [21] discusses how one can use this inter-block information for obtaining additional independent tests of significance for every hypothesis pertaining to the treatments.

Define $Q_i^* = T_i - Q_i - r\bar{y}$, where Q_i is the ith adjusted treatment total, T_i is the total for the ith treatment and $\bar{y}$ is the grand average of all measurements. Then the best estimates of the treatments using both the intra- and inter-block information can be written as

$$\bar{t}_i = \frac{1}{R(k-1)} \{kP_i + d_1 S_1(P_i) + d_2 S_2(P_i)\}, \tag{4.2}$$

where

$$\begin{cases} P_i = WQ_i + W^* Q_i^*, \\ R = r\left[W + \dfrac{W^*}{k-1}\right], \\ W = \dfrac{1}{\sigma^2}, \qquad W^* = \dfrac{1}{\sigma^2 + k\sigma_b^2}. \end{cases} \tag{4.3}$$

The constants d_1, d_2 are usually tabulated with all the designs. Note that (4.2) is the same form as (3.4) except that P_i replaces Q_i, R replaces r, and d_1, d_2 replace c_1, c_2. Thus all results in Section 3 carry over directly by substituting the above changes in the formulas of Section 3 and replacing σ^2 by unity.

On the other hand, under certain conditions which are elaborated in [4], one can also obtain additional independent tests of significance using only the inter-block information. Three cases have to be considered depending on whether the group divisible design is a regular, singular, or semi-regular design. These are the three exhaustive classes of group divisible designs introduced by Bose and Connor [3].

For the regular group divisible designs the inter-block treatment estimates can be written as

$$t_i^* = \frac{kQ_i^* + c_1^* S_1(Q_i^*) + c_2^* S_2(Q_i^*)}{r} \tag{4.4}$$

and will have a variance of

$$\text{Var } t_i^* = \left[\frac{vk - [k + (n-1)c_1^* + n(m-1)c_2^*]}{rv}\right](\sigma^2 + k\sigma_b^2).$$

Also if t_i^* and t_j^* are sth associates ($s = 1, 2$),

$$\text{Cov }(t_i^*, t_j^*) = \left[\frac{vc_s^* - [k + (n-1)c_1^* + n(m-1)c_2^*]}{rv}\right](\sigma^2 + k\sigma_b^2).$$

The quantities c_1^* and c_2^* are defined by

$$c_s^* = \frac{c_s\Delta - r\lambda_s}{\Delta - rH + r^2} \qquad (s = 1, 2),$$

where the parameters c_s, Δ, λ_s, and H are the usual parameters for partially balanced designs (cf. [2], [5]). Therefore all results for Section 3 apply directly by replacing σ^2 by $(\sigma^2 + k\sigma_b^2)$, c_s by c_s^*, and $r(k - 1)$ by r. This results in the two efficiencies being

$$E_a^* = \frac{1}{k - c_1^* + n(c_1^* - c_2^*)},$$

$$E_b^* = \frac{1}{k - c_1^*},$$

and the breakdown of the $v - 1$ treatment sum of squares, using only the inter-block information, is similar to Table 1. If $b > v$, there will be an independent estimate of $\sigma^2 + k\sigma_b^2$, thus permitting independent inter-block tests of significance for the main effects and interactions.

With respect to the singular designs, the intra-block efficiencies are

$$E_a < 1, \qquad E_b = 1.$$

Hence it is only possible to obtain inter-block estimates for those main effects and interactions associated only with the A factors. Since treatments in the same group are first associates, $1/n[t_i + S_1(t_i)]$ represents the average of the group to which treatment i belongs. This average is estimated by

$$\text{(4.5)} \qquad \frac{1}{n}[t_i^* + S_1(t_i^*)] = \frac{Q_i^*}{E_a^* r}, \qquad E_a^* = \frac{mn - k}{k(m - 1)}.$$

There will be m such estimates, thus making it possible to have S-functions of the form $S[1, 2, \cdots, p; 0 \mid i]$ for $p \leqq g$. Then all results of Section 3 follow by replacing E_a by E_a^* and σ^2 by $\sigma^2 + k\sigma_b^2$. If $b > m$, this will permit an estimate of $\sigma^2 + k\sigma_b^2$ and thus we can have independent inter-block tests of significance for the A factor.

The semi-regular group divisible designs have the intra-block efficiencies $E_a = 1$, $E_b < 1$. Therefore it is possible to obtain inter-block estimates of those main effects and interactions having the intra-block efficiency E_b. These $(v - m)$ contrasts can be found by using the normal equations

$$\text{(4.6)} \qquad \sum_{s=1}^{v} \frac{\lambda_{is}}{k} t_s^* = Q_i^*, \qquad i = 1, 2, \cdots, v$$

where $\lambda_{ii} = r$, and λ_{is} = number of blocks in which treatments i and s appear together. The rank of Eqs. (4.6) is exactly $(v - m)$. If $b > (v - m)$, then it will be possible to have an independent estimate of $(\sigma^2 + k\sigma_b^2)$, thus allowing independent inter-block tests of significance for testing these contrasts. An open problem is to simplify this analysis.

Extension to balanced incomplete block designs. Similar results apply to the recovery of inter-block information for the balanced incomplete block designs. The best combined estimate can be written as

$$\bar{t}_i = \frac{P_i}{r[WE + W^*(1 - E)]} = \frac{P_i}{\bar{E}R}, \qquad \bar{E} = \frac{R(k - 1) + \lambda(W - W^*)}{R}.$$

Therefore all results of Section 3 can also be carried over by substituting unity for σ^2, R for r, etc. This produces an efficiency of

$$\bar{E} = \frac{R(k-1) + \lambda(W - W^*)}{R}.$$

In addition if one wished to obtain additional independent significance tests using the inter-block information only, the treatment estimates can be written

$$t_i^* = \frac{Q_i^*}{(1-E)r}$$

and all results of Section 3 follow by replacing $\sigma^2 + k\sigma_b^2$ for σ^2, and

$$E_a = E_b = 1 - E.$$

Again we will have two independent tests of significance for testing every null hypothesis pertaining to the main effects and interactions.

Acknowledgement. The author is indebted to Professor R. C. Bose and Dr. Max Halperin for drawing the author's attention to Lemmas 2.1 and 2.2, respectively. Also thanks go to Dr. Alan Goldman for reading a preliminary version of this paper.

REFERENCES

[1] R. C. Bose, "Mathematical theory of the symmetric factorial design," *Sankhyā*, Vol. 8 (1947), pp 107–166.

[2] R. C. Bose, W. H. Clatworthy, and S. S Shrikhande, *Tables of Partially balanced designs with two associate classes*, Institute of Statistics of Univ of North Carolina, Reprint series No 50 (1954).

[3] R. S. Bose and W. S Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat*, Vol 23 (1952), pp. 367–383

[4] R. C. Bose and K Kishen, "On the problem of confounding in general symmetrical factorial designs," *Sankhyā*, Vol 5 (1940), pp 21–26

[5] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced incomplete block-designs with two associate classes," *J. Amer. Stat. Assn.*, Vol 47 (1952), pp. 151–184.

[6] R. C. Bose, S S. Shrikhande, and K. N. Bhattacharya, "On the construction of group divisible incomplete block designs," *Ann Math. Stat.*, Vol. 24 (1953), pp 167–195.

[7] W. H. Clatworthy, "Partially balanced incomplete block designs with two associate classes and two treatments per block," *Journal of Research of the National Bureau of Standards*, Vol. 54 (1955), pp. 177–190.

[8] W. H. Clatworthy, *Contributions on partially balanced incomplete block designs with two associate classes*, National Bureau of Standards Applied Mathematics Series 47, 1956.

[9] W. G. Cochran and G. M. Cox, *Experimental Designs*, New York, John Wiley and Sons, 1950.

[10] W. T. Federer, *Experimental Designs*, New York, Macmillan, 1955.

[11] R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, Edinburgh.

[12] R. A. Fisher, "The theory of confounding in factorial experiments in relation to the theory of groups," *Ann. Eugenics*, Vol. 11 (1942), pp. 341–353.

[13] O. Kempthorne, *The Design and Analysis of Experiments*, New York, John Wiley and Sons, 1952.

[14] C. Y. KRAMER AND R. A. BRADLEY, "Intra-block analysis for factorials in two-associate class group divisible designs," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 349–361.

[15] J. LI, "Design and statistical analysis of some confounded factorial experiments," *Iowa Agricultural Experimental Station Research Bulletin*, 333, 1934.

[16] K. R. NAIR AND C. R. RAO, "Confounding in asymmetrical factorial experiments," *J. Roy. Stat. Soc.*, Series B, Vol. 10 (1948), pp. 109–131.

[17] C. R. RAO, "General methods of analysis for incomplete block designs," *J. Amer. Stat. Assn.*, Vol. 42 (1947), pp. 541–561.

[18] C. R. RAO, "On the recovery of inter-block information in varietal trials," *Sankhyā*, Vol. 17 (1956), pp. 105–114.

[19] F. YATES, *The design and analysis of factorial experiments*, Imperial Bureau Soil Science, Harpenden, England, 1937.

[20] F. YATES, "The recovery of inter-block information in balanced incomplete block designs," *Ann. Eugenics*, Vol. 10 (1940), pp. 317–325.

[21] M. ZELEN, "The analysis of incomplete block designs," *J. Amer. Stat. Assn.*, Vol. 52 (1957), pp. 204–217.

[22] M. ZELEN, "The use of incomplete block designs for factorial experiments" (abstract), *J. Amer. Stat. Assn.*, Vol. 50 (1955), p. 599.

ACCELERATED STOCHASTIC APPROXIMATION[1]

By Harry Kesten

Cornell University

1. Summary. Using a stochastic approximation procedure $\{X_n\}$, $n = 1, 2, \cdots$, for a value θ, it seems likely that frequent fluctuations in the sign of $(X_n - \theta) - (X_{n-1} - \theta) = X_n - X_{n-1}$ indicate that $|X_n - \theta|$ is small, whereas few fluctuations in the sign of $X_n - X_{n-1}$ indicate that X_n is still far away from θ. In view of this, certain approximation procedures are considered, for which the magnitude of the nth step (i.e., $X_{n+1} - X_n$) depends on the number of changes in sign in $(X_i - X_{i-1})$ for $i = 2, \cdots, n$. In theorems 2 and 3,

$$X_{n+1} - X_n$$

is of the form $b_n Z_n$, where Z_n is a random variable whose conditional expectation, given $X_1, \cdots, X_n$, has the opposite sign of $X_n - \theta$ and b_n is a positive real number. b_n depends in our processes on the changes in sign of

$$X_i - X_{i-1} (i \leqq n)$$

in such a way that more changes in sign give a smaller b_n. Thus the smaller the number of changes in sign before the nth step, the larger we make the correction on X_n at the nth step. These procedures may accelerate the convergence of X_n to θ, when compared to the usual procedures ([3] and [5]). The result that the considered procedures converge with probability one may be useful for finding optimal procedures. Application to the Robbins-Monro procedure (Theorem 2) seems more interesting than application to the Kiefer-Wolfowitz procedure (Theorem 3).

2. *Statement of the theorem.* *The formulation of the theorem is similar to* that of the theorem given by Dvoretzky [2]. Let θ be a real number and

$$T_n (n = 1, 2, \cdots)$$

be measurable transformations. Let X_1 and $Y_n (n = 1, \cdots)$ be random variables[2] and $\{a_n\}$ a sequence of positive numbers and define

$$X_{n+1}(\omega) = T_n(X_1(\omega), \cdots, X_n(\omega)) + b_n(\omega) Y_n(\omega). \quad (1)$$

The sequence $\{b_n(\omega)\}$ is selected in the following way from the sequence $\{a_n\}$

Received March 28, 1957

[1] Note added in proof: The author learned recently that investigation of the above procedure had been suggested by Professor H. Robbins long ago.

[2] X_n, Y_n, and Z_n denote random variables, whereas x_n is used to denote values taken by the random variables

$$
\begin{aligned}
& b_1 = a_1, \\
(2) \qquad & b_2 = a_2, \\
& b_n = a_{t(n)},
\end{aligned}
$$

where

$$(3) \qquad t(n) = 2 + \sum_{i=3}^{n} \iota[(X_i - X_{i-1})(X_{i-1} - X_{i-2})]$$

and

$$\iota(x) = \begin{cases} 1 & \text{if } x \leqq 0 \\ 0 & \text{if } x > 0. \end{cases}$$

Thus, every time $(X_i - X_{i-1})$ differs in sign from $(X_{i-1} - X_{i-2})$ we take another a_n.

Let $\alpha_n(x_1, \cdots, x_n)$, $\beta_n(x_1, \cdots, x_n)$, $\gamma_n(x_1, \cdots, x_n)$ be nonnegative functions and put

$$(4) \qquad \epsilon_N = \sup_{\{x_i\}} \sum_{n=N}^{\infty} \beta_n(x_1, \cdots, x_n),$$

$$(5)^3 \qquad \rho(\delta) = \inf_{n=1,2,\cdots} \inf_{\substack{|x_n - \theta| > \delta \\ x_1, \cdots, x_{n-1} \text{ arbitrary}}} \frac{\gamma_n(x_1, \cdots, x_n)}{b_n}.$$

THEOREM 1. *If*

$$(6) \qquad | T_n(x_1, \cdots, x_n) - \theta | \leqq \begin{cases} (1 + \beta_n(x_1, \cdots, x_n))| x_n - \theta | \\ - \gamma_n(x_1, \cdots, x_n) \textit{ when } (T_n - \theta)(x_n - \theta) > 0 \\ \alpha_n(x_1, \cdots, x_n) \textit{ when } (T_n - \theta)(x_n - \theta) \leqq 0, \end{cases}$$

$$(7) \qquad \lim_{t(n) \to \infty} \alpha_n(x_1, \cdots, x_n) = 0 \qquad \textit{uniformly, for all sequences } x_1, x_2, \cdots \textit{ with } t(n) \to \infty,$$

$$(8)^3 \qquad \lim_{n \to \infty} \frac{(x_n - \theta)\beta_n(x_1, \cdots, x_n)}{b_n} = 0 \qquad \textit{uniformly, for all sequences } x_1, x_2, \cdots,$$

and

$$(9) \qquad \lim_{N \to \infty} \epsilon_N = 0,$$

$$(10) \qquad \rho(\delta) > 0 \qquad \textit{for every positive } \delta,$$

$$(11) \qquad \sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty, \qquad \textit{and} \qquad a_{n+1} \leqq a_n,$$

[3] In (5), (8), and in (13), b_n depends on $x_1, \cdots, x_n$ as given in (2).

(12)[4] $E(Y_n \mid X_1, \cdots, X_n) = 0,$

$$E(Y_n^2 \mid X_1, \cdots, X_n) \leqq \sigma^2 \textit{ with probability } 1,$$

(13)[3]
$$\liminf_{n\to\infty} \lim_{r\to 0} \inf_{\substack{0<|x_n-\theta|\leqq r \\ x_1,\cdots,x_{n-1} \text{ arbitrary}}} \cdot P\{T_n(X_1, \cdots, X_n) + b_n Y_n \geqq X_n \mid X_1 = x_1, \cdots, X_n = x_n\} > 0,$$

$$\liminf_{n\to\infty} \lim_{r\to 0} \inf_{\substack{0<|x_n-\theta|\leqq r \\ x_1,\cdots,x_{n-1} \text{ arbitrary}}} \cdot P\{T_n(X_1, \cdots, X_n) + b_n Y_n < X_n \mid X_1 = x_1, \cdots, X_n = x_n\} > 0,$$

then X_n converges to θ with probability 1.

PROOF OF CONVERGENCE. Without loss of generality we take $\theta = 0$. Also we assume in the following $E \mid X_1 \mid < \infty$. This can be done, because replacing X_1 by

$$X_1^1 = \begin{cases} X_1 & \text{if } \mid X_1 \mid < A \\ A & \text{if } \mid X_1 \mid \geqq A \end{cases}$$

changes the process only with a probability equal to

$$P\{\mid X_1 \mid > A\}.$$

By taking A large enough, this probability becomes arbitrary small. We frequently do not write all the arguments of the functions, e.g., we write β_n for $\beta_n(x_1, \cdots, x_n)$. We shall first prove several lemmas. From

$$E(Y_n \mid X_1, \cdots, X_n) = 0$$

and $E(Y_n^2 \mid X_1, \cdots, X_n) \leqq \sigma^2$ follows immediately.

LEMMA 1. *There exists a function $p(\delta)$ with $0 < p(\delta) < 1$ for $\delta > 0$, and such that*

$$P\left\{Y_n \geqq \frac{\delta}{2} > 0 \mid X_1, \cdots, X_n\right\} \leqq 1 - p(\delta) < 1,$$

$$P\left\{Y_n \leqq -\frac{\delta}{2} < 0 \mid X_1, \cdots, X_n\right\} \leqq 1 - p(\delta) < 1.$$

LEMMA 2.

$$\liminf_{n\to\infty} P\left\{X_{n+1} - X_n \geqq \frac{-\rho^1(\delta)b_n}{2} \middle| X_1, \cdots, X_n; \quad X_n \geqq \delta \text{ and } t(n) \geqq k\right\}$$
$$\leqq 1 - p(\rho^1(\delta)),$$

[4] $P\{\cdot|\cdot\}$ and $E(\cdot|\cdot)$ denote conditional probabilities and conditional expectations respectively.

$$\liminf_{n\to\infty} P\left\{X_{n+1} - X_n \leqq \frac{\rho^1(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n; \quad X_n \leqq -\delta \quad \text{and} \quad t(n) \geqq k\right\} \leqq 1 - p(\rho^1(\delta)),$$

where

$$\rho^1(\delta) = \begin{cases} \rho(\delta) & \text{when} \quad \rho(\delta)a_k \leqq \delta \\ \dfrac{\delta}{a_k} & \text{when} \quad \rho(\delta)a_k > \delta. \end{cases}$$

PROOF. Since $t(n) \geqq k$, we have $b_n \leqq a_k$ and for $X_n \geqq \delta$

$$X_{n+1} \leqq \max[0, (1 + \beta_n)X_n - \gamma_n] + b_n Y_n \leqq X_n + b_n \left[\frac{\beta_n X_n}{b_n} - \rho^1(\delta)\right] + b_n Y_n = X_n + b_n \left[\frac{\beta_n}{b_n} X_n - \rho^1(\delta) + Y_n\right].$$

So by (S) and Lemma 1, we have

$$\liminf_{n\to\infty} P\left\{X_{n+1} - X_n \geqq -\frac{\rho^1(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n; \quad X_n \geqq \delta, \quad t(n) \geqq k\right\}$$
$$\leqq \liminf_{n\to\infty} P\left\{Y_n \geqq \frac{\rho^1(\delta)}{2} - \epsilon \,\middle|\, X_1, \cdots, X_n; \quad X_n \geqq \delta, \quad t(n) \geqq k\right\}$$

for every $\epsilon > 0$.

Application of Lemma 1 gives the first inequality. Similarly we prove the second part of the lemma.

LEMMA 3. *For every k and N*

$$P\{t(n) = k \quad \textit{for} \quad n \geqq N \quad \textit{and} \quad X_n \nrightarrow 0\} = 0$$

(i.e., *when $X_{n+1} - X_n$ does only change sign a finite number of times, then X_n converges to 0*).

PROOF. When $t(n)$ is constant for $n \geqq N$, then X_n is monotonic for $n \geqq N$. Therefore $\{X_n\}$ converges (possibly to $+\infty$ or $-\infty$). Let the limit be positive, say X. But by Lemma 2 for every $\delta > 0$ and $\epsilon < [\rho^1(\delta)a_k]/2$,

$$\lim_{N\to\infty} P\{X_{n+1} - X_n > -\epsilon \quad \text{and} \quad X_n \geqq \delta \quad \text{and} \quad t(n) \geqq k \quad \text{for all } n \geqq N \cdot | \delta \leqq X_N \quad \text{and} \quad t(N) \geqq k\} = 0,$$

so the probability that $X > \delta$ is zero. Similarly the probability $X < -\delta$ is zero. Since δ is an arbitrary positive number, this proves the lemma.

This lemma allows us to limit ourselves in the sequel to those sequences with $t(n) \to \infty$ and therefore $b_n \to 0$.

LEMMA 4. *Let δ be a fixed positive number. Then there exist positive numbers*

n_0 and t_0 such that, whenever $n \geqq n_0$, $t(n) \geqq t_0$ and $|X_n| \geqq \delta$, one has

$$E\{|X_{n+1}| \,|\, X_1, \cdots, X_n; \quad |X_n| \geqq \delta\} \leqq |X_n| - \frac{b_n}{4}\rho(\delta).$$

PROOF. Choose t_0 such that $\alpha_n(X_1, \cdots, X_n) \leqq \delta/2$ for $t(n) \geqq t_0$ and

$$a_{t_0} \leqq \min\left(\frac{2\delta}{4\sigma + \rho(\delta)}, \frac{\delta\rho(\delta)}{16\sigma^2}, \frac{\delta}{\rho(\delta)}\right). \tag{14}$$

Then $b_n \leqq a_{t_0}$ for $t(n) \geqq t_0$. We distinguish two cases

$$|T_n(X_1, \cdots, X_n)| \leqq \frac{\delta}{2}. \tag{a}$$

$$E\left\{|X_{n+1}| \,|\, X_1, \cdots, X_n; |X_n| \geqq \delta, |T_n| \leqq \frac{\delta}{2}\right\}$$

$$\leqq \frac{\delta}{2} + b_n E|Y_n| \leqq \frac{\delta}{2} + b_n\sigma \leqq \delta - \frac{b_n\rho(\delta)}{4} \leqq |X_n| - \frac{b_n\rho(\delta)}{4},$$

$$|T_n(X_1, \cdots, X_n)| > \frac{\delta}{2}. \tag{b}$$

As $\alpha_n(X_1, \cdots, X_n) \leqq \delta/2$ for $t(n) \geqq t_0$, we must have $T_n \cdot X_n > 0$ (cf. (6)). Let $X_n \geqq \delta$. Denote the distribution function of $Y_n(X_1, \cdots, X_n)$ by $H_n(y \mid X)$. As $X_{n+1} = T_n + b_n Y_n$, we have by (12) and (14)

$$E\left\{|X_{n+1}| \,\middle|\, X_1, \cdots, X_n; X_n \geqq \delta, T_n \geqq \frac{\delta}{2}\right\}$$

$$= \int_{-T_n/b_n}^{\infty} (T_n + b_n y)\, dH_n(y \mid X) - \int_{-\infty}^{-T_n/b_n} (T_n + b_n y)\, dH_n(y \mid X)$$

$$\leqq T_n + b_n\left[\int_{-T_n/b_n}^{\infty} y\, dH_n(y \mid X) - \int_{-\infty}^{-T_n/b_n} y\, dH_n(y \mid X)\right]$$

$$= T_n - 2b_n \int_{-\infty}^{-T_n/b_n} y\, dH_n(y \mid X)$$

$$\leqq T_n + 2b_n\left[\int_{-\infty}^{-T_n/b_n} y^2\, dH_n(y \mid X) \int_{-\infty}^{-T_n/b_n} dH_n(y \mid X)\right]^{1/2}$$

$$\leqq T_n + 2b_n\sigma \frac{b_n\sigma}{T_n} \leqq T_n + \frac{4b_n^2\sigma^2}{\delta} \leqq T_n + b_n\frac{\rho(\delta)}{4}.$$

But by (8),

$$|T_n| \leqq |X_n| + b_n\left\{\frac{\beta_n|X_n|}{b_n} - \rho(\delta)\right\} \leqq |X_n| - b_n\frac{\rho(\delta)}{2}$$

for sufficiently large n, say $n \geqq n_0$. For $X_n \leqq -\delta$, the proof is similar. Thus,

in all cases

$$E\left\{|X_{n+1}|\,\middle|\,X_1, \cdots, X_n; |X_n| \geqq \delta\right\} \leqq |X_n| - \frac{b_n\rho(\delta)}{4}.$$

LEMMA 5. *For every* $0 < \delta < \delta' < \delta''$

$$P\{\delta < \liminf |X_n| < \delta' \quad \text{and} \quad \delta'' < \limsup |X_n| \quad \text{and} \quad t(n) \to \infty\} = 0.$$

PROOF. Choose t_0 and n_0, corresponding to δ as introduced in the preceding lemma. Assume now

$$P\{\delta < \liminf |X_n| < \delta' \quad \text{and} \quad \delta'' < \limsup |X_n| \quad \text{and} \quad t(n) \to \infty\} > 0.$$

Then there exist an $n_1 \geqq n_0$ and $t_1 \geqq t_0$ such that

$$(15) \quad P\{\delta < \liminf |X_n| < \delta' \quad \text{and} \quad \delta'' < \limsup |X_n| \quad \text{and} \quad |X_n| > \delta \quad \text{for all} \quad n \geqq n_1 \quad \text{and} \quad b_{n_1} = a_{t_1}\} > 0.$$

Now introduce a new process.

$$Z_i = |X_i| \quad \text{if} \quad i = 1, \cdots, n_1$$

and

$$Z_{n_1+i} = \begin{cases} |X_{n_1+i}| & \text{if} \quad \delta < Z_{n_1+j} \quad \text{for} \quad j = 0, 1, \cdots, i-1 \quad \text{and} \quad b_{n_1} = a_{t_1}, \\ 0 & \text{otherwise.} \end{cases}$$

Unless $b_{n_1} \neq a_{t_1}$ or $|X_j| \leqq \delta$ for some $j \geqq n_1$, we have always $Z_i = |X_i|$, and thus, by (15), also

$$(16) \qquad P\{\delta \leqq \liminf Z_n < \delta' \quad \text{and} \quad \delta'' < \limsup Z_n\} > 0.$$

But by Lemma 4

$$0 \leqq E(Z_{n+1} | X_1, \cdots, X_n) \leqq EZ_n \qquad \text{for} \quad n \geqq n_1.$$

So application of the semimartingale convergence theorem (Loève [4], p. 393) shows that (16) cannot be true. This proves the lemma.

LEMMA 6. $P\{\liminf |X_n| = \infty$ *and* $t(n) \to \infty\} = 0$.

PROOF. If the proposition were not true, we could find, analogous to the last proof, a process Z_i with

$$(17) \quad 0 \leqq E(Z_{n+1} | X_1, \cdots, X_n) \leqq EZ_n \quad \text{for sufficiently large } n, \quad \text{say} \quad n \geqq n_2,$$

and

$$(18) \qquad P\{\liminf Z_n = \infty\} > 0.$$

But as we took $E|X_1| < \infty$, one would have

$$(19) \qquad E|Z_{n_2}| < \infty.$$

However, (17) and (19) together are in contradiction with (18). This proves the lemma.

From Lemmas 3, 5, and 6 one may conclude that with probability 1 either that $\liminf |X_n| = 0$ or $|X_n|$ converges to a finite positive number. We now prove that the last possibility has probability zero.

LEMMA 7.

$$P\{|X_n| \text{ converges to } X \text{ and } 0 < \delta < X < \delta' < \infty \text{ and } t(n) \to \infty\} = 0.$$

PROOF. Choose n_0 and t_0 corresponding to δ, as introduced in Lemma 4. Assume now

$$(20) \quad P\{|X_n| \text{ converges to } X \text{ and } 0 < \delta < X < \delta' < \infty \text{ and } t(n) \to \infty\} > 0.$$

Again there exist an $n_1 \geqq n_0$ and a $t_1 \geqq t_0$ such that

$$(21) \quad P\{\delta < |X_n| < \delta' \text{ for all } n \geqq n_1 \text{ and } b_{n_1} = a_{t_1}\} > 0 \text{ and } a_{t_1}\rho(\delta) \leqq \delta.$$

By Lemma 2 we can choose n_1 and t_1 so that at the same time for $n \geqq n_1$

$$(22) \quad P\left\{X_{n+1} - X_n \geqq -\frac{\rho(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n;\ X_n \geqq \delta,\ t(n) \geqq t_1\right\} \leqq 1 - \frac{p(\rho(\delta))}{2},$$

$$(23) \quad P\left\{X_{n+1} - X_n \leqq \frac{\rho(\delta)b_n}{2} \,\middle|\, X_1, \cdots, X_n;\ X_n \leqq -\delta,\ t(n) \geqq t_1\right\} \leqq 1 - \frac{p(\rho(\delta))}{2}.$$

As before we construct a new process.

$$Z_i = |X_i| \quad \text{if} \quad i = 1, \cdots, n_1$$

and

$$Z_{n_1+i} = \begin{cases} |X_{n_1+i}| \text{ if } \delta < Z_{n_1+j} < \delta' & \text{for } j = 0, \cdots, i-1 \text{ and } b_{n_1} = a_{t_1}, \\ Z_{n_1+i-1} - a_{t_1+i-1}\dfrac{\rho(\delta)}{4} & \text{otherwise.} \end{cases}$$

From (21) follows

$$(24) \qquad P\{\delta < Z_n < \delta' \text{ for all } n \geqq n_1\} > 0,$$

and thus,

$$(25) \qquad P\left\{\left|\sum_{k=n_1}^{n} (Z_{k+1} - Z_k)\right| < 2(\delta' - \delta) \text{ for all } n \geqq n_1\right\} > 0.$$

Denote

$$E(Z_{k+1} - Z_k \mid X_1, \cdots, X_k) \quad \text{by} \quad m_k(X_1, \cdots, X_k) \qquad (= m_k \text{ for short}).$$

By Lemma 4 and the construction of the Z-process,

$$m_k(X_1, \cdots, X_k) \leqq -\frac{c_k}{4}\rho(\delta) \text{ for } k \geqq n_1, \tag{26}$$

where

$$c_{n_1+i} = \begin{cases} b_{n_1+i} & \text{if} \quad \delta < Z_j < \delta' \quad \text{for} \quad j = 0, \cdots, i \quad \text{and} \quad b_n = a_{t_1}, \\ a_{n_1+i} & \text{otherwise.} \end{cases}$$

Further for $k \geqq n_1$,

$$\operatorname{var}(Z_{k+1} - Z_k \mid X_1, \cdots, X_k) \leqq c_k^2\left[\frac{\rho^2(\delta)}{16} + EY_k^2\right] \leqq c_k^2 C, \tag{27}$$

where

$$C = \frac{\rho^2(\delta)}{16} + \sigma^2.$$

In addition

$$\sum_{k=n_1}^{n} c_k \geqq \sum_{k=0}^{n-n_1} a_{t_1+k}. \tag{28}$$

By (25),

$$P\left\{\left|\sum_{k=n_1}^{n}(Z_{k+1} - Z_k - m_k)\right| \geqq \left|\sum_{k=n_1}^{n} m_k\right| - 2(\delta' - \delta) \qquad \text{for all } n \geqq n_1\right\} > 0, \tag{29}$$

and thus, by (26) and (28),

$$P\left\{\left|\sum_{k=n_1}^{n}(Z_{k+1} - Z_k - m_k)\right| \geqq \frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta) \qquad \text{for all } n \geqq n_1\right\} > 0. \tag{30}$$

But for

$$\frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta) > 0$$

we have by Tchebycheff's inequality and (27)

$$P\left\{\left|\sum_{k=n_1}^{n}(Z_{k+1}-Z_k-m_k)\right| \geqq \frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k}-2(\delta'-\delta)\right\}$$

(32)
$$\leqq \frac{C\sum_{k=n_1}^{n} Ec_k^2}{\left\{\frac{\rho(\delta)}{4}\sum_{k=0}^{n-n_1} a_{t_1+k}-2(\delta'-\delta)\right\}^2},$$

(32)
$$\sum_{k=n_1}^{n} Ec_k^2 \leqq \sum_{k=0}^{n-n_1} a_{t_1+k}^2 Er_{t_1+k},$$

where

$$r_{t_1+k} = \text{number of times } c_{n_1+i} = a_{t_1+k}.$$

As soon as the Z_i process differs from the $|X_i|$ process, we don't keep the same a_{t_1+k} for more than one step. Therefore $Er_{t_1+k} \leqq 1 +$ expected number of times that $\{c_{n_1+i} = a_{t_1+k}$ and $\delta < Z_{n_1+j} < \delta'$ for $j = 1, \cdots, i$ and $b_{n_1} = a_{t_1}\}$ occurs.

If $\delta < X_{n_1+i} < \delta'$, then by (22),

$$P\{X_{n_1+i+1} > X_{n_1+i} \mid \delta < |X_{n_1+j}| < \delta' \quad j = 1, \cdots, i \text{ and } b_{n_1} = a_{t_1} \text{ and } X_{n_1+i} > \delta\} \leqq 1 - \frac{p(\rho(\delta))}{2},$$

and

$$P\left\{\begin{matrix} X_{n_1+i+l} > X_{n_1+i+l-1} \mid \delta < |X_{n_1+j}| < \delta' \quad j = 1, \cdots, i \\ l = 1, \cdots, s \end{matrix} \text{ and } b_{n_1} = a_{t_1} \text{ and } X_{n_1+i} > \delta\right\} \leqq \left\{1 - \frac{p(\rho(\delta))}{2}\right\}^s.$$

As we pick a new a_j as soon as $(X_i - X_{i-1})(X_{i-1} - X_{i-2}) \leqq 0$, we have: Expected number of times that

$$\{c_{n_1+i} = a_{t_1+k} \quad \text{and} \quad \delta < Z_{n_1+j} < \delta' \quad j = 1, \cdots, i \quad \text{and} \quad b_{n_1} = a_{t_1}\}$$

under the condition that $X_{j+1} \geqq X_j > \delta$ for the first $j \geqq n_1$ with $b_j = a_{t_1+k}$, is at most

(33)
$$2 + \left(1 - \frac{p(\rho(\delta))}{2}\right) + \left(1 - \frac{p(\rho(\delta))}{2}\right)^2 + \cdots \leqq \frac{3}{p(\rho(\delta))}.$$

The case where $X_{j+1} < X_j$ at the first time that $b_j = a_{t_1+k}$ is more difficult. Let us divide the interval (δ, δ') in

$$\left[\frac{2(\delta'-\delta)}{\rho(\delta)a_{t_1+k}}\right] + 1$$

non-overlapping intervals[5] I_t with

[5] $[a]$ is the largest integer $\leqq a$.

$$\text{length } (I_t) < \frac{\rho(\delta)a_{t_1+k}}{2}\left(t = 1, 2, \cdots, \left[\frac{2(\delta' - \delta)}{\rho(\delta)a_{t_1+k}}\right] + 1\right).$$

Expected number of times that

$$\{c_{n_1+i} = a_{t_1+k} \quad \text{and} \quad \delta < Z_{n_1+j} < \delta' \; j = 1, \cdots, i, \; Z_{\tilde{n}_1+i} \,\varepsilon\, I_t\}$$

under the condition that $X_{j+1} < X_j$, $X_j > \delta$ for first $j \geqq n_1$, with $b_j = a_{t_1+k}$, is at most

$$1 + \left(1 - \frac{p(\rho(\delta))}{2}\right) + \left(1 - \frac{p(\rho(\delta))}{2}\right)^2 \cdots = \frac{2}{p(\rho(\delta))}.$$

This can be proved analogous to (33) using (22) and the fact that

$$\text{length } (I_t) < [\rho(\delta)a_{t_1+k}]/2.$$

As there are

$$\left[\frac{2(\delta' - \delta)}{\rho(\delta)a_{t_1+k}}\right] + 1$$

intervals I_t, expected number of times that

$$\{c_{n_1+i} = a_{t_1+k} \quad \text{and} \quad \delta < Z_{n_1+j} < \delta' \; j = 1, \cdots, i, \quad \text{and} \quad b_{n_1} = a_{t_1}\}$$

under the condition that $X_{j+1} < X_j$, $X_j > \delta$ for the first $j \geqq n_1$ with

$$b_j = a_{t_1+k},$$

is at most

$$\frac{2\left\{\left[\frac{2(\delta' - \delta)}{\rho(\delta)a_{t_1+k}}\right] + 1\right\}}{p(\rho(\delta))}.$$

Similar estimates are valid when $X_j < -\delta$ for the first $j \geqq n_1$ with $b_j = a_{t_1+k}$.

As $a_{t_1+k} \to 0 (k \to \infty)$, we can find a positive constant D such that

$$Er_{t_1+k} \leqq \frac{D}{a_{t_1+k}}.$$

By (31) and (32), it follows that

$$P\left\{\left|\sum_{n=n_1}^{n} (Z_{k+1} - Z_k - m_k)\right| \geqq \frac{\rho(\delta)}{4} \sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta)\right\}$$

(34)

$$\leqq \frac{CD \sum_{k=0}^{n-n_1} a_{t_1+k}}{\left\{\frac{\rho(\delta)}{4} \sum_{k=0}^{n-n_1} a_{t_1+k} - 2(\delta' - \delta)\right\}^2}.$$

As $\sum_{k=0}^{\infty} a_{t_1+k} = \infty$, the right hand side of (34) tends to zero when $n \to \infty$

and therefore (29) cannot be true and

$$P\{|X_n| \text{ converges to } X \neq 0 \quad \text{and} \quad t(n) \to \infty\} = 0.$$

Combining the remark after Lemma 6, and Lemma 7 we proved

$$P\{\liminf |X_n| = 0\} = 1. \tag{35}$$

Until now we only used that a_n tends monotonically to zero and

$$\textstyle\sum_{n=1}^{\infty} a_n = \infty,$$

but not yet $\sum_{n=1}^{\infty} a_n^2 < \infty$.

LEMMA 8. *Define*

$$s(n) = \begin{cases} 1 & \text{if} \quad T_n(X_1, \cdots, X_n)\cdot X_n > 0 \\ -1 & \text{if} \quad T_n(X_1, \cdots, X_n)\cdot X_n \leqq 0, \end{cases}$$

$$Y_n^1 = Y_n \prod_{j=1}^{n} s(j),$$

$$d(m, m-1) = 1$$

$$d(m, n) = \prod_{j=m}^{n} (1 + \beta_j) \qquad (n \geqq m),$$

$$S(m+1, n) = \sum_{j=m}^{n} d(j+1, n) b_j Y_j^1.$$

Then the conditions

$$\alpha_{m+j-1}(X_1, \cdots, X_{m+j-1}) \leqq \frac{\epsilon}{8} \qquad j = 1, \cdots, k,$$

$$d(m, \infty) \leqq \frac{3}{2},$$

$$|X_m| \leqq \frac{\epsilon}{4},$$

$$|X_{m+j}| > \frac{\epsilon}{4} \qquad j = 1, \cdots, k-1,$$

and

$$\sup_{n \geqq m} |S(m+1, n)| \leqq \frac{\epsilon}{16}$$

imply

$$|X_{m+j}| \leqq \frac{\epsilon}{2} \qquad j = 1, \cdots, k.$$

The proof follows immediately from Wolfowitz [6], p. 1154. We need the following

COROLLARY. *If $t(m)$ is so large that*

$$\alpha_{m+j-1}(X_1, \cdots, X_{m+j-1}) \leqq \frac{\epsilon}{8} \qquad j = 1, 2, \cdots$$

and if

$$d(m, \infty) \leqq \frac{3}{2},$$

then

$$P\left\{ |X_{m+j}| > \frac{\epsilon}{2} \text{ for some positive integer } j \,\Big|\, |X_m| \leqq \frac{\epsilon}{4}\right\}$$
$$\leqq P\left\{ \sup_{n_1, n_2 \geqq m} |S(n_1 + 1, n_2)| \geqq \frac{\epsilon}{16}\right\}$$
$$\leqq \frac{32^2}{\epsilon^2} \sum_{j=m}^{\infty} \operatorname{var}\{d(j+1, n) b_j Y_j^1\} \leqq \left(\frac{48\sigma}{\epsilon}\right)^2 \sum_{j=m}^{\infty} Eb_j^2.$$

PROOF OF THEOREM 1. In view of Lemma 3, we only have to prove

$$P\{\limsup |X_n| > 0 \quad \text{and} \quad t(n) \to \infty\} = 0.$$

By condition (13)

$$2\zeta = \min\Big(\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{0<|x_n|\leqq\tau} P\{X_{n+1} - X_n \geqq 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\},$$
$$\liminf_{n\to\infty} \lim_{\tau\to 0} \inf_{0<|x_n|\leqq\tau} P\{X_{n+1} - X_n < 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\}\Big) > 0$$

Take $\xi > 0$ and n_2 such that

$$\begin{aligned} &P\{X_{n+1} - X_n \geqq 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\} > \zeta > 0, \\ &P\{X_{n+1} - X_n < 0 \mid X_1, \cdots, X_{n-1}, X_n = x_n\} > \zeta > 0 \end{aligned} \tag{36}$$

for $0 < |x_n| \leqq \xi$ and $n \geqq n_2$.

Choose an $\epsilon \leqq \xi$ and t_2 such that

$$\alpha_n(X_1, \cdots, X_n) \leqq \frac{\epsilon}{8} \quad \text{when} \quad t(n) \geqq t_2,$$

and let

$$d(n_2, \infty) \leqq \frac{3}{2}.$$

Let now for some $m \geqq n_2$

$$|X_m| \leqq \frac{\epsilon}{4} \quad \text{and} \quad t(m) \geqq t_2.$$

Construct the following process

$$Z_k^{(m)} = X_k \quad \text{if} \quad k = 1, \cdots, m$$

and

$$Z_{m+i}^{(m)} = T_{m+i-1}(Z_1, \cdots, Z_{m+i-1}) + c_{m+i-1}Y_{m+i-1}(X_1, \cdots, X_{m+i-1})(i = 1, 2, \cdots),$$

where the c's are determined in the following way:

$$c_m = b_m = a_{t(m)}$$

(37) $$c_{m+i} = \begin{cases} c_{m+i-1} & \text{if } |Z_{m+j}| \leqq \frac{\epsilon}{2} \quad j = 0, 1, \cdots, i \\ & \text{and } (Z_{m+i} - Z_{m+i-1})(Z_{m+i-1} - Z_{m+i-2}) > 0 \\ a_l & \text{if } |Z_{m+j}| \leqq \frac{\epsilon}{2} \quad j = 0, 1, \cdots, i \\ & \text{and } (Z_{m+i} - Z_{m+i-1})(Z_{m+i-1} - Z_{m+i-2}) \leqq 0 \\ & \text{and } c_{m+i-1} = a_{l-1} \\ a_{t(m)+i} & \text{otherwise.} \end{cases}$$

Then Er_l = expected number of times $c_{m+j} = al$ is zero when $l < t(m)$. For $l \geqq t(m)$ it is at most

(38) $$1 + (1 - \zeta) + (1 - \zeta)^2 \cdots = \frac{1}{\zeta}.$$

In fact from (36) and (37),

$$P\{c_{m+j} = c_{m+j-1}\} \leqq 1 - \zeta.$$

Using (38) and applying the corollary of Lemma 8 to the $Z(m)$ process, and thus replacing the b's by the c's, one finds for $m \geqq n_2$,

$$P\left\{|X_{m+j}| > \frac{\epsilon}{2} \text{ for some positive integer } j \,\middle|\, |X_m| \leqq \frac{\epsilon}{4}, t(m) \geqq t_2\right\}$$

$$\leqq P\left\{|Z_{m+j}^{(m)}| > \frac{\epsilon}{2} \text{ for some positive integer } j \,\middle|\, |Z_m^{(m)}| \leqq \frac{\epsilon}{4}, t(m) \geqq t_2\right\}$$

$$\leqq \left(\frac{48\sigma}{\epsilon}\right)^2 \sum_{n=t_2}^{\infty} \frac{a_n^2}{\zeta}.$$

Now choose $t_3 \geqq t_2$ such that

$$\left(\frac{48\sigma}{\epsilon}\right)^2 \frac{1}{\zeta} \sum_{n=t_3}^{\infty} a_n^2 \leqq \frac{\epsilon}{2},$$

and $n_3 \geqq n_2$ such that

$$P\left\{\left[\,|X_n| > \frac{\epsilon}{4} \quad \text{for all} \quad n \geqq n_3 \quad \text{or} \quad t(n_3) < t_3\right] \quad \text{and} \quad t(n) \to \infty\right\} \leqq \frac{\epsilon}{2}$$

(such an n_3 exists by (35)). Then

$$P\{\limsup |X_n| > \epsilon \quad \text{and} \quad t(n) \to \infty\}$$

$$\leqq P\left\{\left[\,|X_n| > \frac{\epsilon}{4} \quad \text{for all} \quad n \geqq n_3 \quad \text{or} \quad t(n_3) < t_3\right] \quad \text{and} \quad t(n) \to \infty\right\}$$

$$+ \sum_{m=n_3}^{\infty} P\left\{X_m \quad \text{is the first after} \quad X_{n_3-1} \quad \text{with} \quad |X_m| \leqq \frac{\epsilon}{4}\right.$$

$$\left.\text{and} \quad t(n_3) \geqq t_3 \quad \text{and} \quad \max_{k \geqq m} |Z_k(m)| > \frac{\epsilon}{2}\right\}$$

$$\leqq \frac{\epsilon}{2} + \frac{\epsilon}{2} \sum_{m=n_3}^{\infty} P\left\{X_m \quad \text{is the first after} \quad X_{n_3-1} \quad \text{with} \quad |X_m| \leqq \frac{\epsilon}{4}\right\} \leqq \epsilon.$$

As the only restriction on ϵ is $\epsilon \leqq \xi$, this proves the theorem.

3. Applications.

Accelerated Robbins-Monro procedure.

THEOREM 2. *Let X_1 and $Y(x)$ be random variables and $\{a_n\}$ a sequence of positive numbers and define*

$$X_{n+1}(\omega) = X_n(\omega) - b_n(M(X_n) - \alpha) + b_n Y(X_n).$$

The sequence $\{b_n\}$ is selected in the following way from the sequence $\{a_n\}$:

$$b_1 = a_1,$$

$$b_2 = a_2,$$

$$b_n = a_{t(n)},$$

(cf. (2) and (3)).

If $M(x)$ is a measurable function satisfying

(39) $$(x - \theta)(M(x) - \alpha) > 0 \qquad \text{for } x \neq \theta,$$

(40) $$\inf_{\delta \leqq |x-\theta| < \infty} |M(x) - \alpha| > 0 \qquad \text{for every} \quad \delta > 0,$$

(41) $|M(x) - \alpha| \leqq c + d\,|x - \theta|$ for some positive constants c and d,

and if

(42) $$\sum_{n=1}^{\infty} a_n = \infty, \qquad \sum_{n=1}^{\infty} a_n^2 < \infty, \quad \text{and} \quad a_{n+1} \leqq a_n,$$

(43) $$E(Y(X_n)|\,X_1, \cdots, X_n) = 0, \qquad E(Y^2(X_n)|\,X_1, \cdots, X_n) \leqq \sigma^2$$

with probability 1,

$$\lim_{\tau\to 0} \inf_{0<|x-\theta|\leq\tau} P\{Y(x) - M(x) + \alpha \geqq 0\} > 0 \tag{44}$$
$$\lim_{\tau\to 0} \inf_{0<|x-\theta|\leq\tau} P\{Y(x) - M(x) + \alpha < 0\} > 0,$$

then

$$P\{X_n \text{ converges to } \theta\} = 1.$$

PROOF. Take

$$\alpha_n = \begin{cases} b_n(c + d\,|\,x_n - \theta\,|) & \text{for} \quad b_n d > 1 \\ b_n c & \text{for} \quad b_n d \leqq 1, \end{cases}$$

$$\beta_n \equiv 0,$$

$$\gamma_n = b_n\,|\,M(x_n) - \alpha\,|$$

in Theorem 1.

The process as described in Theorem 2 gives a stochastic approximation method for the point θ which uses the number of changes in sign in

$$(X_i - X_{i-1})(X_{i-1} - X_{i-2}) \qquad i = 3, \cdots, n$$

to determine $(X_{n+1} - X_n)$. We only reduce b_n and thus the magnitude of

$$X_{n+1} - X_n$$

when the last two corrections $X_n - X_{n-1}$ and $X_{n-1} - X_{n-2}$ had different signs. As indicated in the summary this process may pull X_n to θ faster (for large $|\,X_n - \theta\,|$) than the Robbins-Monro procedure. In Theorem 2 the conditions are slightly stronger than for the Robbins-Monro process as given by Blum [1]. Blum does not need

$$a_{n+1} \leqq a_n$$

or (44) and has

$$\inf_{\delta\leq|x-\theta|\leq\delta'} |\,M(x) - \alpha\,| > 0 \quad \text{for every} \quad 0 < \delta \leqq \delta' < \infty \tag{40a}$$

instead of (40).

One can easily give an example to show that we cannot replace (40) by (40a) and the following example shows that (44) cannot be dispensed with.

EXAMPLE. Take

$$\theta = 0, \qquad \alpha = 0, \qquad a_n = \frac{1}{n}.$$

Let $\{x_{2n+1,0}\}(n = 0, 1, \cdots)$ be a sequence of real numbers such that

$$x_{2n+1,0} \neq x_{2m+1,0} \quad \text{for} \quad n \neq m \quad \text{and} \quad 1 \leqq x_{2n+1,0} \leqq 2.$$

Let $\{x_{2n,0}\}(n = 1, 2, \cdots)$ be a sequence of real numbers such that

$$x_{2n,0} \neq x_{2m,0} \quad \text{for} \quad n \neq m \quad \text{and} \quad -2 \leqq x_{2n,0} \leqq -1.$$

We now construct recursively sequences $\{x_{n,k}\}(k = 0, 1, \cdots)$. Put

$$Z(x) = M(x) - Y(x)$$

and

$$Z_{n,k} = Z(x_{n,k}),$$

so

$$X_{n+1} = X_n - b_n Z(X_n).$$

We start with $\{x_{1,k}\}$ by taking

$$Z_{1,0} = \begin{cases} z'_{1,0} = x_{1,0} - x_{2,0} & \text{with probability } \frac{1}{2} \\ z''_{1,0} & \text{with probability } \frac{1}{2}, \end{cases}$$

where $\frac{1}{2}x_{1,0} < z''_{1,0} < x_{1,0}$. Further take $x_{1,1} = x_{1,0} - z''_{1,0}$, and in general

$$Z_{1,k} = \begin{cases} z'_{1,k} = x_{1,k} - x_{2,0} & \text{with probability } \frac{1}{2} \\ z''_{1,k} & \text{with probability } \frac{1}{2} \end{cases}$$

and

$$x_{1,k+1} = x_{1,k} - z''_{1,k},$$

where

$$\tfrac{1}{2}x_{1,k} < z''_{1,k} < x_{1,k}.$$

For $n > 1$ we take

$$Z_{n,0} = \begin{cases} z'_{n,0} = (n-1)(x_{n,0} - x_{n+1,0}) & \text{with probability } \dfrac{1}{2n^2} \\ z''_{n,0} & \text{with probability } 1 - \dfrac{1}{2n^2}, \end{cases}$$

$$x_{n,1} = x_{n,0} - \frac{1}{n-1} z''_{n,0},$$

where $z''_{n,0}$ is such that

$$z''_{n,0} \cdot x_{n,0} > 0 \quad \text{and} \quad \tfrac{1}{2}|x_{n,0}| < |z''_{n,0}| < |x_{n,0}|$$

and $x_{n,1}$ is not equal to any $x_{m,i}$ with $m < n$. Further for $k > 0$,

$$Z_{n,k} = \begin{cases} z'_{n,k} = n(x_{n,k} - x_{n+1,0}) & \text{with probability } \dfrac{1}{2n^2} \\ z''_{n,k} & \text{with probability } 1 - \dfrac{1}{2n^2}, \end{cases}$$

$$x_{n,k+1} = x_{n,k} - \frac{1}{n} z''_{n,k},$$

where $z''_{n,k}$ is such that

$$z''_{n,k} \cdot x_{n,k} > 0, \qquad \tfrac{1}{2} | x_{n,k} | < | z''_{n,k} | < | x_{n,k} |$$

and $x_{n,k+1}$ is not equal to any $x_{m,s}$ with $m < n$. We take $M(x_{n,k}) = EZ_{n,k}$ and

$$Y(x_{n,k}) = Z_{n,k} - M(x_{n,k}).$$

For $x \neq x_{n,k}$ for all n, k, we take $M(x)$ and $Y(x)$ in any way such that the conditions of Theorem 2, except (44), are satisfied.

Take now $X_1 = x_{1,0}$ with probability 1. By the choice of $z'_{n,k}$, we get the value $x_{n+1,0}$ as soon as $Z_{n,k}$ takes the value $z'_{n,k}$. But for every n, with probability 1, Z will take once the value $z'_{n,k}$. Therefore with probability 1, all the values $x_{n,0}$ occur in the sequence $\{X_n\}$ and thus,

$$P\{X_n \text{ converges to } 0\} = 0.$$

Accelerated Kiefer-Wolfowitz procedure.

THEOREM 3. *Let X_1 and $Y(x)$ be random variables and let $\{a_n\}$ be a sequence of positive numbers and u some positive constant and define*

$$X_{n+1}(\omega) = X_n(\omega) - b_n[M(X_n - u) - M(X_n + u)] + b_n[Y(X_n - u) - Y(X_n + u)].$$

The sequence $\{b_n\}$ is selected in the following way from the sequence $\{a_n\}$:

$$b_1 = a_1,$$
$$b_2 = a_2,$$
$$b_n = a_{t(n)},$$

(cf. (2) and (3)). *If $M(x)$ is a measurable function, satisfying*

$$\text{(45)} \qquad \begin{aligned} &\inf_{x-\theta \geq \delta} \{M(x-u) - M(x+u)\} > 0 \\ &\inf_{x-\theta \leq -\delta} \{M(x-u) - M(x+u)\} < 0 \end{aligned}$$

for every $\delta > 0$,

$$\text{(46)} \qquad | M(x-u) - M(x+u) | \leq c + d\, | x - \theta |$$

for some positive constants c and d, and if

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty, \quad \text{and} \quad a_{n+1} \leqq a_n, \tag{47}$$

$$\begin{aligned} E(Y(X_n - u) - Y(X_n + u) \mid X_1, \cdots, X_n) &= 0, \\ E((Y(X_n - u) - Y(X_n + u))^2 \mid X_1, \cdots, X_n) &\leqq \sigma^2 \end{aligned} \tag{48}$$

with probability 1,

$$\begin{aligned} &\lim_{\tau \to 0} \inf_{0<|x-\theta| \leqq \tau} P\{Y(x - u) - Y(x + u) - M(x - u) + M(x + u) \\ &\qquad\qquad \geqq 0\} > 0 \\ &\lim_{\tau \to 0} \inf_{0<|x-\theta| \leqq \tau} P\{Y(x - u) - Y(x + u) - M(x - u) + M(x + u) \\ &\qquad\qquad < 0\} > 0 \end{aligned} \tag{49}$$

then

$$P\{X_n \text{ converges to } \theta\} = 1.$$

PROOF. Take

$$\alpha_n = \begin{cases} b_n(c + d \mid x_n - \theta \mid) & \text{for} \quad b_n d > 1 \\ b_n c & \text{for} \quad b_n d \leqq 1, \end{cases}$$

$$\beta_n \equiv 0,$$

$$\gamma_n = b_n \mid M(x_n - u) - M(x_n + u) \mid$$

in Theorem 1.

REMARK. Theorem 3 is also implied by Theorem 2. The procedure in Theorem 3 requires u to be independent of n, and therefore differs from the usual Kiefer Wolfowitz procedure ([3]). Also condition (45) does not imply that $M(x)$ has a maximum, or if it has one, that θ is the location of the maximum. However for every y with $| y - \theta | > u$, there exists an x with $| x - \theta | \leqq u$, such that $M(x) > M(y)$.

Acknowledgement. The author wishes to thank Professor J. Wolfowitz for suggesting the possibility of generalizing Theorem 2 and for his many other helpful and critical remarks.

REFERENCES

[1] J. R. BLUM, "Approximation methods which converge with probability one," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 382–386.

[2] A. Dvoretzky, "On stochastic approximation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, 1956, pp. 39–55.
[3] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 462–466.
[4] M. Loève, *Probability Theory*, D. van Nostrand, 1955.
[5] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400–407.
[6] J. Wolfowitz, "On stochastic approximation methods," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1151–1156.

TESTING THE HYPOTHESIS THAT TWO POPULATIONS DIFFER ONLY IN LOCATION[1]

By Balkrishna V. Sukhatme

Indian Council of Agricultural Research, New Delhi

0. Summary. Let $X_1, X_2, \cdots, X_n$ be n independent identically distributed random variables with cumulative distribution function $F(x - \xi)$. Let

$$\hat{\xi}(X_1, X_2, \cdots, X_n)$$

be an estimate of ξ such that $\sqrt{n}(\hat{\xi} - \xi)$ is bounded in probability. The first part of this paper (Secs. 2 through 4) is concerned with the asymptotic behavior of U-statistics modified by centering the observations at $\hat{\xi}$. A set of necessary and sufficient conditions are given under which the modified U-statistics have the same asymptotic normal distribution as the original U-statistics. These results are extended to generalized U-statistics and to functions of several generalized U-statistics. The second part gives an application of the asymptotic theory developed earlier to the problem of testing the hypothesis that two populations differ only in location.

1. Introduction. Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be two independent samples of observations from populations with cumulative distribution functions $F(x - \xi)$ and $G(x - \eta) = F[(x - \eta)/\delta]$ respectively, ξ and η being the unknown location parameters and δ a scale parameter. No knowledge is assumed concerning the distribution functions F and G except that they are absolutely continuous. The problem considered in this paper is that of testing the hypothesis that the two populations differ only in location against the alternative that the Y's are more spread out than the X's and vice versa, or in symbols

$$(1.1) \qquad H: \delta = 1, \qquad A: \delta \neq 1.$$

From intuitive considerations and the work of Fraser [1], it seems likely that there do not exist similar tests for testing the hypothesis H, which are very satisfactory. The following simplified problem was therefore considered by the author [2]. Let the location parameters ξ and η be known, say $\xi = \eta = 0$, so that the distribution functions of X and Y differ only in the scale parameter. Then the problem considered is that of testing the hypothesis

$$H': \delta = 1, \quad \text{i.e., } F = G,$$
$$A: \delta \neq 1, \quad \text{i.e., } F \neq G.$$

Received May 18, 1956.

[1] This research was performed while the author was at the Statistical Laboratory, University of California, Berkeley, and was supported by the Office of Ordnance Research, U.S. Army, under contract DA-04-200-ORD-171.

Several nonparametric tests have been suggested for testing the hypothesis H', particularly by Mood [3]. The author [2] has considered some of these tests and discussed their asymptotic properties from the point of view of power considerations. These tests are based on what are known as generalised U-statistics and are reasonably efficient. But our main interest lies not in testing the hypothesis H' but H. However, once we have a class $\{W_N\}$ of tests for testing the hypothesis H', a class of tests $\{\hat{W}_N\}$ for testing the hypothesis H suggests itself. This class of tests may be obtained as follows. We obtain suitable estimates of the parameters ξ and η and then apply any of the tests of the class W_N to the deviations of the X's and the Y's from the respective estimates.

If the X's and the Y's come from normal populations, the usual test of significance for testing the hypothesis H is the variance ratio test based on the statistic

$$(1.2)\qquad F = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{\sum_{i=1}^{m} (X_i - \bar{X})^2} \cdot \frac{m-1}{n-1},$$

which is also the most commonly used statistical test for comparing sample variances. Usually, however, since little is known about the populations from which the samples are drawn, this test is used as if the assumption of normality could be ignored. It appears, however, that this is not the case.

The sensitivity to non-normality of the F-test was first pointed out by E. S. Pearson [4] whose findings were later confirmed by Geary [5] and Gayen [6]. They showed that the F-test is particularly sensitive to changes in Kurtosis from the normal theory value of zero. It is easy to see that the F-statistic when suitably normalised is asymptotically distribution free. More recently, Box and Andersen ([7] and [8]) have studied this problem in great detail and have shown on the basis of extensive sampling experiments that the F-statistic so normalized is insensitive to departures from normality.

Since the tests considered in [2] are nonparametric and reasonable for normal alternatives, it appears that they might be more efficient for non-normal alternatives and also more stable for small samples. We propose, therefore, to investigate whether such tests, after modification by the introduction of estimates of parameters are asymptotically distribution free.

This is achieved by considering the asymptotic theory of generalised U-statistics modified by the introduction of estimates of parameters, which is given in Secs. 3 and 4. In Sec. 5, it is shown that the nonparametric test proposed in [2], after modification, is asymptotically distribution free for populations with bounded and symmetric probability densities. It turns out however that even under such restrictive conditions, the nonparametric test proposed by Mood, after modification is not asymptotically distribution free. Finally, the last section considers the small sample behavior of the proposed test for some particular alternatives.

2. Some definitions and known results.

DEFINITION 2.1. Let X_{ij}, $j = 1, 2, \cdots, n_i$ for a fixed i be independent random variables identically distributed with c.d.f. $F_i(x)$ and density function $f_i(x)$. Let i run from 1 to k and $s_1 \leqq n_1$, $s_2 \leqq n_2, \cdots, s_k \leqq n_k$. Further, let

$$\varphi(u_1, \cdots, u_{s_1}; v_1, \cdots, v_{s_2}; \cdots; w_1, w_2, \cdots, w_{s_k})$$

be a function symmetric in each set of its arguments. Then the statistic

$$U_N = \binom{n_1}{s_1}^{-1} \binom{n_2}{s_2}^{-1} \cdots \binom{n_k}{s_k}^{-1}$$

$$\cdot \sum \varphi(X_{1,\alpha_1}, \cdots, X_{1,\alpha s_1}; X_{2,\beta_1}, \cdots, X_{2,\beta s_2}; \cdots; X_{k,\delta_1}, \cdots X_{k,\delta s_k}),$$

where the summation runs over all subscripts α, β, δ such that

$$1 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_{s_1} \leqq n_1$$

$$1 \leqq \beta_1 < \beta_2 < \cdots < \beta_{s_2} \leqq n_2$$

$$\cdots$$

$$1 \leqq \delta_1 < \delta_2 < \cdots < \delta_{s_k} \leqq n_k$$

is called a generaliesd U-statistic.

Let $\rho_1, \rho_2, \cdots, \rho_k$ be k fixed numbers such that $n_i = N\rho_i$ and $\sum_{i=1}^{k} \rho_i = 1$. Then Lehmann [9] has shown that $\sqrt{N}[U_N - EU_N]$ is asymptotically normally distributed with mean zero and asymptotic variance σ^2 given by

$$\sigma^2 = \frac{s_1^2}{\rho_1} \zeta_{100\cdots0} + \frac{s_2^2}{\rho_2} \zeta_{010\cdots0} + \cdots + \frac{s_k^2}{\rho_k} \zeta_{00\cdots01},$$

where

$$\zeta_{00\cdots1\cdots0} = E\varphi_1\varphi_2 - [E\varphi_1]^2,$$

1 occurs at the ith place in $\zeta_{00\cdots1\cdots0}$,

$$\varphi_1 = \varphi(X_{11}, \cdots, X_{1s_1}; \cdots; X_{i1}, X_{i2}, \cdots, X_{is_i}; \cdots)$$

and φ_2 is obtained from φ_1 by replacing all the X_{jk} by X'_{jk} excepting X_{i1}, the primes denoting a new set of independent random variables. This result is a generalisation of the U-statistics considered by Hoeffding [10].

For the sake of simplicity, we shall restrict ourselves to the two sample problem only. The extension to k samples is straight forward.

DEFINITION 2.2. As before, let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be two independent samples drawn from populations with c.d.f.'s $F(x - \xi)$ and $G(x - \eta)$ respectively. Further, let

$\hat{\xi}(X_1, \cdots, X_m)$ and $\hat{\eta}(Y_1, \cdots, Y_n)$ be estimates of ξ

and η, the two location parameters. Then the generalised U-statistic with the observations centered at the respective location parameters and the modified generalised U-statistic for the two sample problem are respectively,

$$U_N = \binom{m}{s_1}^{-1}\binom{n}{s_2}^{-1}\sum_{\alpha,\beta}\varphi(X_{\alpha_1}-\xi, \cdots, X_{\alpha_{s_1}}-\xi; Y_{\beta_1}-\eta, \cdots, Y_{\beta_{s_2}}-\eta),$$

$$\hat{U}_N = \binom{m}{s_1}^{-1}\binom{n}{s_2}^{-1}\sum_{\alpha,\beta}\varphi(X_{\alpha_1}-\hat{\xi}, \cdots, X_{\alpha_{s_1}}-\hat{\xi}; Y_{\beta_1}-\hat{\eta}, \cdots, Y_{\beta_{s_2}}-\hat{\eta}),$$

where $\varphi(u_1, \cdots, u_{s_1}; v_1, \cdots, v_{s_2})$ is a function symmetric in u and in v and the summation runs over all subscripts α, β such that

$$1 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_{s_1} \leqq m,$$

$$1 \leqq \beta_1 < \beta_2 < \cdots < \beta_{s_2} \leqq n.$$

DEFINITION 2.3. Let $\hat{W}_N$ be a test based on the statistic $\hat{U}_N$. If the asymptotic distribution of $\hat{U}_N$ is independent of the original populations from which the X's and the Y's are drawn under the null hypothesis, the test $\hat{W}_N$ will be said to be asymptotically distribution free.

Finally we define a quantity L_N required in the study of the asymptotic behavior of modified generalised U statistics.

DEFINITION 2.4.

$$L_N = \binom{m}{s_1}^{-1}\binom{n}{s_2}^{-1}$$

$$\cdot \textstyle\sum_{\alpha,\beta}[\varphi(X_{\alpha_1}-\hat{\xi}, \cdots, X_{\alpha_{s_1}}-\hat{\xi}; Y_{\beta_1}-\hat{\eta}, \cdots, Y_{\beta_{s_2}}-\hat{\eta}) - A(\hat{\xi}-\xi, \hat{\eta}-\eta)],$$

where

$$A(t_1-\xi, t_2-\eta) = E\varphi(X_1-t_1, \cdots, X_{s_1}-t_1; Y_1-t_2, \cdots, Y_{s_2}-t_2),$$

expectation being taken with respect to all the X's and the Y's.

3. The limiting distribution of L_N. In this section, we will prove theorems, giving the conditions under which L_N and U_N have the same asymptotic normal distribution. We will start with one sample problem and then extend the result to two samples. In what follows, we write $\mathcal{L}(X_n) \to \mathcal{L}(X)$ (read: the distribution law of X_n converges to the distribution law of X), or $\lim_{n\to\infty} \mathcal{L}(X_n) = \mathcal{L}(X)$ if $F_n(a) \to F(a)$ at every point a of continuity of F where F_n and F are the c.d.f.'s of X_n and X, respectively.

THEOREM 3.1. *Let $X_1, X_2, \cdots, X_n$ be n independent identically distributed*

random variables with c.d.f. $F(x - \xi)$. *Let* $\varphi(u_1, u_2, \cdots, u_s)$ *with* $s \leqq n$ *be a real valued symmetric function of its arguments such that if*

$$(3.0) \qquad W(x_1, x_2, \cdots, x_s, t) = \varphi(x_1 - t, \cdots, x_s - t) - A(t - \xi),$$

where $A(t - \xi) = E\varphi(X_1 - t, \cdots, X_s - t)$, *the following conditions are satisfied.*

$$(\mathrm{B}_1) \qquad |W(x_1, x_2, \cdots, x_s, t)| \leqq M_1, \text{ and } E|W(X_1, \cdots, X_s; t + h) - W(X_1, \cdots, X_s; t)| \leqq M_2 h,$$ *M_1 and M_2 being fixed constants.*

There exists a sequence $\{t_j\}$ *such that for each set of x's*

$$(\mathrm{B}_2) \qquad \sup_{0 \leq t_j \leq k} | W(x_1, \cdots, x_s, t_j) - W(x_1, \cdots, x_s, 0) | = \sup_{0 \leq t \leq k} | W(x_1, \cdots, x_s, t) - W(x_1, \cdots, x_s, 0)|.$$

Further, let $\hat{\xi}(X_1, X_2, \cdots, X_n)$ *be an estimate of* ξ *such that given* $\Sigma_1 > 0$, *there exists a number b such that for n sufficiently large*

$$(3.1) \qquad P\left\{| \hat{\xi} - \xi | \geqq \frac{b}{\sqrt{n}}\right\} \leqq \Sigma_1.$$

Define

$$(3.2) \qquad U_n = \binom{n}{s}^{-1} \sum \varphi(X_{\alpha_1} - \xi, \cdots, X_{\alpha_s} - \xi),$$

the summation being taken over all subscripts α such that

$$1 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_s \leqq n$$

and

$$L_n = \binom{n}{s}^{-1} \sum [\varphi(X_{\alpha_1} - \hat{\xi}, \cdots, X_{\alpha_s} - \hat{\xi}) - A(\hat{\xi} - \xi)].$$

Then

$$(3.3) \qquad \lim_{n \to \infty} \mathfrak{L}(\sqrt{n} L_n) = \lim_{n \to \infty} \mathfrak{L}(\sqrt{n} [U_n - EU_n]) = N(0, s^2 \zeta_1),$$

where

$$(3.4) \qquad \zeta_1 = E\varphi_1^2(X_1 - \xi) - E^2\varphi(X_1 - \xi, \cdots, X_s - \xi),$$

$$\varphi_1(x_1 - \xi) = E\varphi(x_1 - \xi, X_2 - \xi, \cdots, X_s - \xi).$$

PROOF. For the sake of simplicity we may, without loss of generality, assume $\xi = 0$. However, before we proceed further, we shall first prove the following lemma, which we shall use in the proof of the theorem.

LEMMA 3.2. *Let*

$$H_{r,n}(x_1, x_2, \cdots, x_s, t) = \sup_{\frac{r\delta}{\sqrt{n}} \leqq z \leqq t} \left| W(x_1, \cdots, x_s, z) - W\left(x_1, \cdots, x_s, \frac{r\delta}{\sqrt{n}}\right) \right| \tag{3.5}$$

and

$$S_{r,n}(t) = \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} \left[W\left(x_{\alpha_1}, \cdots, x_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(x_{\alpha_1}, \cdots, x_{\alpha_s}, \frac{\delta r}{\sqrt{n}}\right) \right]. \tag{3.6}$$

Then, if $r\delta \leqq t \leqq (r+1)\delta$ and n is sufficiently large,

$$\text{(i)}\quad EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) \leqq \frac{M_2\delta}{\sqrt{n}}; \tag{3.7}$$

$$\text{(ii)}\quad E\left\{\left[H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) - EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right] \cdot \left[H_{r,n}\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) - EH_{r,n}\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right]\right\} \to 0 \text{ as } n \to \infty, \tag{3.8}$$

where

$$1 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_s \leqq n,$$

$$1 \leqq \beta_1 < \beta_2 < \cdots < \beta_s \leqq n;$$

$$\text{(iii)}\quad E\mid S_{r,n}(t)\mid^2 \leqq \frac{d(t - r\delta)}{\sqrt{n}}, \tag{3.9}$$

where d is a fixed constant and higher powers of $1/\sqrt{n}$ are neglected.

PROOF. (i) and (ii) are easily obtained as consequences of conditions (B_1)

and (B_2) of Theorem 3.1. To prove (iii) we have

$$E \mid S_{r,n}(t) \mid^2 = n \binom{n}{s}^{-2} \sum_{\alpha,\beta} E\left\{\left[W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{r\delta}{\sqrt{n}}\right)\right] \cdot \left[W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{r\delta}{\sqrt{n}}\right)\right]\right\}.$$

Consider a typical term; with c integers common to the two terms. We then have

$$E\left\{\left[W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{r\delta}{\sqrt{n}}\right)\right] \cdot \left[W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{r\delta}{\sqrt{n}}\right)\right]\right\}$$

$$\leqq E\left|\left[W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{r\delta}{\sqrt{n}}\right)\right] \cdot \left[W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{r\delta}{\sqrt{n}}\right)\right]\right|$$

$$\leqq 2M_1 E\left| W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{r\delta}{\sqrt{n}}\right)\right|$$

$$\leqq 2M_1 M_2 \frac{(t - r\delta)}{\sqrt{n}}.$$

The total contribution of such terms to

$$E|S_{r,n}(t)|^2 \leqq n \binom{n}{s}^{-2} \binom{n}{2s - c} \cdot A(t - r\delta) / \sqrt{n},$$

A being some fixed constant. It follows that

$$E \mid S_{r,n}(t) \mid^2 \sim \frac{1}{n^{c-\frac{1}{2}}} (t - r\delta).$$

When $c = 0$, the expectation of the product is zero. Retaining only powers of $1/\sqrt{n}$, the result now follows. Q.E.D.

Proof of Theorem 3.1. Let

$$S_n^{(t)} = \sqrt{n} \binom{n}{s}^{-1} \sum_{\alpha} \left[W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) - W\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, 0\right)\right].$$

Then it is easily seen that

$$S_n(t) = S_{r,n}(t) + S_{0,n}(r\delta).$$



Let $\epsilon > 0$, $\delta = \epsilon/2M_2$, and t be such that $r\delta \leqq t \leqq (r+1)\delta$. Then it is seen that

$$
\begin{aligned}
| S_{r,n}(t) | &\leqq \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{t}{\sqrt{n}}\right) \\
&\leqq \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) \\
&\leqq \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} \left[H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right. \\
&\qquad \left. - EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right] \\
&\qquad + \sqrt{n}EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) \\
&\leqq \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} \left| H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right. \\
&\qquad \left. - EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right| + M_2\delta \\
&= D_1 + M_2\delta,
\end{aligned}
$$

where

$$
\begin{aligned}
D_1 = \sqrt{n}\binom{n}{s}^{-1} \sum_{\alpha} \left[H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right. \\
\left. - EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right].
\end{aligned}
$$

Now

$$
\begin{aligned}
ED_1^2 = n\binom{n}{s}^{-2} \sum_{\alpha,\beta} E\left\{\left[H_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right.\right. \\
\left. - EH_{r,n}\left(X_{\alpha_1}, \cdots, X_{\alpha_s}, \frac{(r+1)\delta}{\sqrt{n}}\right)\right] \\
\cdot \left.\left[H_{r,n}\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{(r+1)\delta}{\sqrt{n}}\right) - EH_{r,n}\left(X_{\beta_1}, \cdots, X_{\beta_s}, \frac{\delta(r+1)}{\sqrt{n}}\right)\right]\right\},
\end{aligned}
$$

the summation having the same meaning as before. Considering again a typical term with c integers common to the two terms, we find that the total contribu-

tion of such terms to ED_1^2 is

$$\le n\binom{n}{s}^{-2}\binom{n}{2s-r} E\left[H_{r,s}\left(X_{\alpha_1},\cdots,X_{\alpha_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right.$$

$$\left.- EH_{r,s}\left(X_{\alpha_1},\cdots,X_{\alpha_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right]$$

$$\cdot\left[H_{r,s}\left(X_{\beta_1},\cdots,X_{\beta_s},\frac{(r+1)\delta}{\sqrt{n}}\right) - EH_{r,s}\left(X_{\beta_1},\cdots,X_{\beta_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right]$$

$$\sim \frac{A}{n^{r-1}} E\left[H_{r,s}\left(X_{\alpha_1},\cdots,X_{\alpha_s},\frac{(r+1)\delta}{\sqrt{n}}\right) - EH_{r,s}\left(X_{\alpha_1},\cdots,X_{\alpha_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right]$$

$$\cdot\left[H_{r,s}\left(X_{\beta_1},\cdots,X_{\beta_s},\frac{(r+1)\delta}{\sqrt{n}}\right) - EH_{r,s}\left(X_{\beta_1},\cdots,X_{\beta_s},\frac{(r+1)\delta}{\sqrt{n}}\right)\right],$$

which tends to zero by Lemma 3.2 for $r \ge 1$. When $r = 0$, the expectation of the product is identically equal to zero.

Hence $ED_1^2 \to 0$ as $n \to \infty$. It follows that

$$P\left\{\sup_{r\delta \le t \le (r+1)\delta} S_n(t)^2 \ge \epsilon_1\right\} \to 0,$$

for every r.

Also $E\, S_{0,n}(r\delta)^2 \le 2M_1M_2 r\delta/\sqrt{n} \to 0$ as $n \to \infty$, therefore,

$$S_n(r\delta) \xrightarrow{P} 0.$$

It follows that $\sup_{t \in C} |S_n(t)| \to 0$, C being some bounded set. Hence,

$$S_n(\sqrt{n}\,\xi) \xrightarrow{P} 0,$$

that is,

$$\sqrt{n}\,L_n - \sqrt{n}\,[U_n - EU_n] \xrightarrow{P} 0,$$

therefore,

$$\lim_{n\to\infty} \mathcal{L}(\sqrt{n}\,L_n) = \lim_{n\to\infty} \mathcal{L}(\sqrt{n}\,[U_n - EU_n]).$$

But by Hoeffding's Theorem 7.1, page 305 of [10], U_n is asymptotically normally distributed, whence the required result follows. Q.E.D.

We complete this section by stating without proof the generalization of the above result to the two sample problem. The proof goes more or less along the same lines as that of Theorem 3.1.

THEOREM 3.3. *Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be two independent samples drawn from populations with c.d.f.'s $F(x - \xi)$ and $G(x - \eta)$ respectively. Further, let $\varphi(u_1, \cdots, u_{s_1}; v_1, \cdots, v_{s_2})$ with $s_1 \le m$ and $s_2 \le n$ be a real-valued*

function symmetric in u and in v separately such that if

$$(3.10)\quad \begin{aligned} &W(x_1, x_2, \cdots, x_{s_1}, y_1, \cdots, y_{s_2}, t_1, t_2) \\ &= \varphi(x_1 - t_1, \cdots, x_{s_1} - t_1; y_1 - t_2, \cdots, y_{s_2} - t_2) - A(t_1 - \xi, t_2 - \eta), \end{aligned}$$

the following conditions are satisfied:

$$(B_3)\quad \begin{aligned} &|W(x_1, x_2, \cdots, x_{s_1}, y_1, y_2, \cdots, y_{s_2}, t_1, t_2| \leqq M_{11} \\ &E|W(X_1, \cdots, X_{s_1}, Y_1, \cdots, Y_{s_2}, t_1 + h, t_2) \\ &\qquad - W(X_1, \cdots, X_{s_1}, Y_1, \cdots Y_{s_2}, t_1, t_2)| \leqq M_{21}h \\ &E|W(X_1, \cdots X_{s_1}, Y_1, \cdots, Y_{s_2}, t_1, t_2 + k) \\ &\qquad - W(X_1, \cdots, X_{s_1}, Y_1, \cdots Y_{s_2}, t_1, t_2)| \leqq M_{22}k, \end{aligned}$$

where M_{11}, M_{21} and M_{22} are certain fixed constants

There exist sequences $\{t_j\}$ and $\{l_j\}$ such that for every set of x's and y's,

$$(B_4)\quad \begin{aligned} &\sup_{\substack{0 \leqq t_j \leqq k_1 \\ 0 \leqq l_j \leqq k_2}} |W(x_1, \cdots, x_{s_1}, y_1, \cdots, y_{s_2}, t_j, l_j) \\ &\qquad - W(x_1, x_{s_1}; y_1, \cdots, y_{s_2}, 0, 0)| \\ &= \sup_{\substack{0 \leqq t \leqq k_1 \\ 0 \leqq l \leqq k_2}} |W(x_1, \cdots, x_{s_1}, y_1, \cdots, y_{s_2}, t, l) \\ &\qquad - W(x_1, \cdots, x_{s_1}; y_1, \cdots, y_{s_2}, 0, 0)|\cdot \end{aligned}$$

Further, let $\hat{\xi}(X_1, \cdots X_m)$ and $\hat{\eta}(Y_1, \cdots, Y_n)$ be estimates of ξ and η respectively such that given $\epsilon_1 > 0$ and $\epsilon_2 > 0$, there exist numbers b_1 and b_2 such that for m and n sufficiently large

$$(3.11)\quad P\left\{ |\hat{\xi} - \xi| \geqq \frac{b_1}{\sqrt{n}} \right\} \leqq \epsilon_1,$$

$$(3.12)\quad P\left\{ |\hat{\eta} - \eta| \geqq \frac{b_2}{\sqrt{n}} \right\} \leqq \epsilon_2.$$

Define

$$(3.13)\quad U_N = \binom{m}{s_1}^{-1} \binom{n}{s_2}^{-1} \sum_{\alpha,\beta} \varphi(X_{\alpha_1} - \xi, \cdots, X_{\alpha_{s_1}} - \xi; Y_{\beta_1} - \eta, \cdots, Y_{\beta_{s_2}} - \eta),$$

the summation being taken over all subscripts α, β such that

$$(3.14)\quad \begin{aligned} &1 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_{s_1} \leqq m, \\ &1 \leqq \beta_1 < \beta_2 < \cdots < \beta_{s_2} \leqq n. \end{aligned}$$

Then if $m = N\rho$ *and* $n = N(1 - \rho)$,

$$\lim_{N\to\infty} \mathfrak{L}(\sqrt{N}\, L_N) = \lim_{N\to\infty} \mathfrak{L}(\sqrt{N}\,(U_N - EU_N)) = N(0, \sigma^2), \tag{3.15}$$

where σ^2 *is the asymptotic variance of* U_N *and is given by*

$$\sigma^2 = \frac{s_1^2}{\rho}\zeta_{10} + \frac{s_2^2}{1-\rho}\zeta_{01}, \tag{3.16}$$

where ζ_{10} *and* ζ_{01} *have the same meaning as in* (2.1).

4. The asymptotic distribution of modified generalised U-statistics. We are now in a position to consider the statistic $\hat{U}_N$ and obtain conditions under which it has the same asymptotic normal distribution as the statistic U_N. This result is contained in Theorem 4.1.

THEOREM 4.1. *If in addition to the conditions of Theorem* 3.1,

(i) $\sqrt{n}(\hat{\xi} - \xi)$ *has a limiting distribution*

and

(ii) $A(t) = E[\varphi(X_1 - t, \cdots, X_s - t) \mid \xi = 0]$ *has a derivative continuous in the neighbourhood of the origin, then*

(a) *If* $A'(0) = 0$, *where* $A'(t) = \frac{d}{dt}A(t)$,

$$\lim_{n\to\infty} \mathfrak{L}(\sqrt{n}\,[\hat{U}_n - EU_n]) = \lim_{n\to\infty} \mathfrak{L}(\sqrt{n}\,[U_n - EU_n]) = N(0, s^2\zeta_1).$$

(b) *If* $A'(0) \neq 0$, $\hat{\xi}$ *is asymptotically normally distributed and the joint distribution of* $\hat{\xi}$ *and* U_n *is asymptotically normal, then* $\sqrt{n}(\hat{U}_n - EU_n)$ *is asymptotically normally distributed.*

PROOF. We have

$$\sqrt{n}[\hat{U}_n - EU_n] = \sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)] + \sqrt{n}[A(\hat{\xi} - \xi) - EU_n].$$

But $A(\hat{\xi} - \xi) = A(0) + (\hat{\xi} - \xi)A'(h)$ where $h = \Delta(\hat{\xi} - \xi)$, $|\Delta| < 1$. Therefore

$$\sqrt{n}[\hat{U}_n - EU_n] = \sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)] + \sqrt{n}(\hat{\xi} - \xi)\cdot A'(h).$$

Since $\sqrt{n}(\hat{\xi} - \xi)$ has a limiting distribution and $A'(0) = 0$, it follows from the continuity considerations and Slutsky's theorem that $\sqrt{n}[\hat{U}_n - EU_n]$ and $\sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)]$ have the same asymptotic distribution. But by Theorem 3.1, $\sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)]$ and $\sqrt{n}[U_n - EU_n]$ have the same asymptotic normal distribution. It follows that $\sqrt{n}[\hat{U}_n - EU_n]$ and $\sqrt{n}[U_n - EU_n]$ have the same asymptotic normal distribution. This proves (a).

To prove (b), it is sufficient to remark that because of Theorem 3.1 and Slutsky's Theorem, the joint distribution of $\sqrt{n}(\hat{\xi} - \xi)$ and $\sqrt{n}[\hat{U}_n - A(\hat{\xi} - \xi)]$ is asymptotically normal. Q.E.D.

In the preceding theorem we make the following observations.
(1) If $A'(0) = 0$, then $\sigma^2(\hat{U}_n) = \sigma^2(U_n)$.
(2) If $A'(0) \neq 0$, then $\sigma^2(\hat{U}_n) = \sigma^2(U_n)$, if and only if

$$A'(0) = \frac{-2\sigma(U_n, \xi)}{\sigma^2(\xi)},$$

where $\sigma(U_n, \xi)$ is the asymptotic covariance between U_n and ξ and $\sigma^2(\xi)$ is the asymptotic variance of ξ.

For the sake of simplicity we will now consider the special case when $s = 1$, ξ is the sample median and $f(x)$ is symmetric about the median which may be taken to be the origin.

Now

$$\begin{aligned} A(t) &= E\varphi(X - t) \\ &= \int \varphi(x - t)f(x)\,dx \\ &= \int \varphi(y)f(y + t)\,dy. \end{aligned}$$

If there exists an integrable function $g(y)$ such that

$$\left| \frac{f(y + t) - f(y + t_0)}{t - t_0} \right| \leqq g(y) \tag{4.1}$$

and the derivative of f exists almost everywhere except for a set of measure zero, then

$$A'(0) = \int \varphi(y)f'(y)\,dy. \tag{4.2}$$

Also it has been shown in [11] that the joint distribution of U_n and ξ is asymptotically normal and that

$$\sigma^2(\xi) = \frac{1}{4nf^2(0)} \tag{4.3}$$

and

$$\sigma(U_n, \xi) = \frac{1}{2nf(0)} \int_0^\infty [\varphi(x) - \varphi(-x)]f(x)\,dx. \tag{4.4}$$

Hence $\sigma^2(\hat{U}_n) = \sigma^2(U_n)$ if and only if

$$\int_0^\infty \left[4f(0) + \frac{f'(x)}{f(x)}\right] [\varphi(x) - \varphi(-x)]f(x)\,dx \underset{f}{\equiv} 0. \tag{4.5}$$

We will now show that the condition (4.5) implies that $\varphi(x) - \varphi(-x) = 0$ almost everywhere. To show this, it is enough to consider the subfamily of

probability densities given by

(4.6) $$f(x, \theta) = \frac{1}{2\theta} e^{-|x|/\theta}.$$

We observe that the derivative of f exists everywhere except at the origin. Also, we have

(4.7) $$\left| \frac{e^{-|x+h|/\theta} - e^{-|x|/\theta}}{h/\theta} \right| \leqq ce^{-|x|/\theta},$$

for h sufficiently small, c being a fixed constant. Condition (4.1) is thus satisfied for the family of distributions (4.6). On substitution, condition (4.5) becomes

(4.8) $$\int_0^\infty e^{-x/\theta}[\varphi(x) - \varphi(-x)]\, dx \underset{\theta}{\equiv} 0,$$

whence it follows from the unicity of the unilateral Laplace transform that $\varphi(x) - \varphi(-x) = 0$ almost everywhere, in which case $A'(0) = 0$ and condition (2) reduces to condition (1).

It is now clear that $A'(0) = 0$ is a necessary and sufficient condition that $\hat{U}_n$ and U_n have the same asymptotic normal distribution.

We will now extend the results of Theorem 4.1 to the two-sample problem.

THEOREM 4.2. *If in addition to the conditions of Theorem* 3.3,

(i) $\sqrt{N}(\hat{\xi} - \xi)$ *and* $\sqrt{N}(\hat{\eta} - \eta)$ *have limiting distributions*

and

(ii) $A(t_1, t_2) = E[\varphi(X_1 - t_1, \cdots, X_{s_1} - t_1, Y_1 - t_2, \cdots, Y_{s_2} - t_2) \mid \xi = \eta = 0]$

possesses first order partial derivatives continuous in the neighborhood of the origin, then

(a) *If*

$$\left.\frac{\partial A(t_1, t_2)}{\partial t_1}\right|_{t_1=t_2=0} = \left.\frac{\partial A(t_1, t_2)}{\partial t_2}\right|_{t_1=t_2=0} = 0,$$

$$\lim_{N\to\infty} \mathfrak{L}(\sqrt{N}(\hat{U}_N - EU_N) = \lim_{N\to\infty} \mathfrak{L}(\sqrt{N}[U_N - EU_N])$$
$$= N(0, \sigma^2),$$

where σ^2 is the asymptotic variance of U_N.

(b) If the above condition is not satisfied, $\hat{\xi}$ and $\hat{\eta}$ are asymptotically normally distributed and the joint distribution of $\hat{\xi}$, $\hat{\eta}$ and the U statistic is asymptotically normal, then $\sqrt{N}[\hat{U}_N - EU_N]$ is asymptotically normally distributed.

PROOF. The proof of this theorem goes in exactly the same lines as that of Theorem 4.1 and is fairly obvious. Q.E.D.

It may be remarked here that the results of Secs. 3 and 4 can be extended to

random vectors as also to functions of several U-statistics. The proof follows in exactly the same way as the theorem on the asymptotic distribution of a function of moments follows from the fact of their asymptotic normality [12]. We shall content ourselves by stating an analogue of Theorem 4.2 as applied to several U-statistics.

THEOREM 4.3. *With reference to the two sample problem, let*

$$\varphi(u_1, u_2, \cdots, u_{s_1(\gamma)}; \text{for}, v_1, v_2, \cdots, v_{s_2(\gamma)}), \qquad \gamma = 1, \cdots, g,$$

with $s_1(\gamma) \leqq m$ *and* $s_2(\gamma) \leqq n$ *be g real valued functions symmetric in u and in v. Further, let*

$$W^{(\gamma)}(x_{\alpha_1}, \cdots, x_{\alpha_{s_1(\gamma)}}; y_{\beta_1}, \cdots, y_{\beta_{s_2(\gamma)}}, t_1, t_2) = \varphi^{(\gamma)}(x_{\alpha_1}, \cdots, x_{\alpha_{s_1(\gamma)}}; y_{\beta_1}, \cdots, y_{\beta_{s_2(\gamma)}}) - A^{(\gamma)}(t_1, t_2),$$

where

$$A^{(\gamma)}(t_1, t_2) = E[\varphi^{(\gamma)}(X_{\alpha_1} - t_1, \cdots, X_{\alpha_{s_1(\gamma)}} - t_1, Y_{\beta_1} - t_2, \cdots, Y_{\beta_{s_2(\gamma)}} - t_2) \mid \xi = \eta = 0]$$

possess partial derivatives continuous in the neighborhood of the origin and $W^{(\gamma)}$ *satisfy the conditions* (B_3) *and* (B_4) *of Theorem* 3.3 *for* $\gamma = 1, \cdot \cdot, g$. *Also let* $\sqrt{N}(\hat{\xi} - \xi)$ *and* $\sqrt{N}(\hat{\eta} - \eta)$ *have limiting distributions where the estimates* $\hat{\xi}$ *and* $\hat{\eta}$ *satisfy the conditions* (3.11) *and* (3.12) *of Theorem* 3.3. *Define*

$$U_N^{(\gamma)} = \binom{m}{s_1(\gamma)}^{-1} \binom{n}{s_2(\gamma)}^{-1} \cdot \sum_{\alpha, \beta} \varphi^{(\gamma)}(X_{\alpha_1} - \xi, \cdots, X_{\alpha_{s_1(\gamma)}} - \xi, Y_{\beta_1} - \eta, \cdots, Y_{\beta_{s_2(\gamma)}} - \eta),$$

the summation having the same meaning as before. Then

(i) *a necessary and sufficient condition that the joint asymptotic distribution of*

$$\sqrt{N}(\hat{U}_N^{(1)} - EU_N^{(1)}), \qquad \cdots, \qquad \sqrt{N}(\hat{U}_N^{(g)} - EU_N^{(g)})$$

be the same as the joint asymptotic distribution of

$$\sqrt{N}(U_N^{(1)} - EU_N^{(1)}), \qquad \cdots, \qquad \sqrt{N}(U_N^{(g)} - EU_N^{(g)})$$

is that

$$\left.\frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_1}\right|_{t_1 = t_2 = 0} = \left.\frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_2}\right|_{t_1 = t_2 = 0} = 0$$

for $\gamma = 1, 2, \cdots, g$.

(ii) *A necessary and sufficient condition that the asymptotic distribution of* $\sqrt{N} \sum_{\gamma=1}^{g} C_\gamma[\hat{U}_N^{(\gamma)} - EU_N^{(\gamma)}]$ *be the same as the asymptotic distribution of*

$$\sqrt{N} \sum_{\gamma=1}^{g} C_\gamma(U_N^{(\gamma)} - EU_N^{(\gamma)})$$

is that

$$\sum_{\gamma=1}^{g} C_\gamma \frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_1}\bigg|_{t_1=t_2=0} = 0$$

and

$$\sum_{\gamma=1}^{g} C_\gamma \frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_2}\bigg|_{t_1=t_2=0} = 0.$$

5. Consequences of Theorem 4.2. In this section, we will consider some of the tests of a class $\{\hat{W}_N\}$ based on a class of statistics $\{\hat{U}_N\}$ for testing the hypothesis that two populations differ only in location and investigate whether they are asymptotically distribution free.

Consider first the test statistic T proposed in [2] based on a sample of m X's and n Y's. The test statistics may be defined as

$$T = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} K(x_i, Y_j), \tag{5.1}$$

where

$$K(X, Y) = 1 \quad \text{if} \begin{cases} \text{either } 0 < X < Y, \\ \text{or } \quad Y < X < 0, \end{cases} \tag{5.2}$$
$$= 0 \quad \text{otherwise.}$$

A corresponding modified test is then based on the statistic

$$\hat{T} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} K(X_i - \tilde{X}, Y_j - \tilde{Y}), \tag{5.3}$$

$\tilde{X}$ and $\tilde{Y}$ being the sample medians. Let $\xi = \eta = 0$. We then have

$$\begin{aligned} A(t_1, t_2) &= EK(X - t_1, Y - t_2) \\ &= \int_0^\infty [1 - G(x + t_2)]\, dF(x + t_1) + \int_{-\infty}^0 G(x + t_2)\, dF(x + t_1). \end{aligned} \tag{5.4}$$

Also $W(x, y, t_1, t_2) = K(x - t_1, y - t_2) - A(t_1, t_2)$. It can then be shown that

$$\begin{aligned} E \mid W(X, Y, t_1, 0) - W(X, Y, 0, 0)| &\leqq 3 \int_{-\infty}^0 | F(x + t_1) - F(x) \,|\, dG(x) + 2 \,|\, F(t_1) - F(0)| \\ &\leqq 5at_1 \end{aligned}$$

if the distribution function F has a derivative F' bounded in absolute value by a. Similarly, it can be shown that

$$E|W(X, Y, 0, t_2) - W(X, Y, 0, 0)| \leqq 5bt_2,$$

provided the distribution function G has a derivative G' bounded in absolute value by b.

Hence the condition (B_3) of Theorem 3.3 is satisfied. Observing that K can be expressed as a difference of two monotone functions, it is easy to see that condition (B_4) is also satisfied. Again, we have

$$\frac{\partial A(t_1, t_2)}{\partial t_1} = -f(t_1)[2G(t_2) - 1] + \int_0^\infty f(x + t_1)\, dG(x + t_2) - \int_{-\infty}^0 f(x + t_1)\, dG(x + t_2),$$

$$\frac{\partial A(t_1, t_2)}{\partial t_2} = -\int_0^\infty g(x + t_2)\, dF(x + t_1) + \int_{-\infty}^0 g(x + t_2)\, dF(x + t_1).$$

Clearly,

$$\left.\frac{\partial A(t_1, t_2)}{\partial t_1}\right|_{t_1 = t_2 = 0} = \left.\frac{\partial A(t_1, t_2)}{\partial t_2}\right|_{t_1 = t_2 = 0} = 0$$

if $f(x)$ and $g(x)$ are symmetric about the origin. Conditions (a) of Theorem 4.2 are satisfied. Hence $\hat{T}$ has the same asymptotic normal distribution as the statistic T. This consequence is stated in Theorem 5 1.

THEOREM 5.1. *If the X's and the Y's are distributed symmetrically about the respective medians and have bounded density functions, the test of the hypothesis H based on the statistic $\hat{T}$ is asymptotically distribution free.*

Consider now the test statistic suggested by Mood [3]. The test statistic may be defined as

$$M = \sum_{i=1}^{n} \left(r_i - \frac{m + n + 1}{2} \right)^2, \tag{5.5}$$

where r_i is the rank of Y_i in the combined sample of $(m + n)$ observations. Noting that

$$r_i = 1 + \sum_{j=1}^{m} \varphi(X_j, Y_i) + \sum_{k=1}^{n} \varphi(Y_k, Y_i), \tag{5.6}$$

where

$$\varphi(u, v) = 1 \quad \text{if} \quad u < v,$$
$$= 0 \quad \text{otherwise},$$

it is easy to see that if $m + n = N$, and

$$\psi(u, v, w) = 1 \quad \text{if} \quad u < w \quad \text{and} \quad v < w,$$
$$= 0 \quad \text{otherwise},$$

$$\frac{M}{N^3} = C_1 U_N^{(1)} + C_2 U_N^{(2)} + C_3 U_N^{(3)} + P\left(\frac{1}{N}\right), \tag{5.7}$$

where C_1, C_2, C_3 are certain known fixed constants. $P(1/N)$ is a third-degree

polynomial in $1/N$ and

$$U_N^{(1)} = \binom{m}{2}^{-1} \binom{n}{1}^{-1} \sum_i^n \sum_{j \neq k}^m \psi(X_j, X_k, Y_i),$$

(5.8)
$$U_N^{(2)} = \binom{m}{1}^{-1} \binom{n}{2}^{-1} \sum_j^m \sum_{k \neq i}^n \psi(X_j, Y_k, Y_i),$$

$$U_N^{(3)} = \binom{m}{1}^{-1} \binom{n}{1}^{-1} \sum_i \sum_j \varphi(X_j, Y_i)$$

are three generalised U-statistics so that

(5.9)
$$\frac{\hat{M}}{N^3} = C_1 \hat{U}_N^{(1)} + C_2 \hat{U}_N^{(2)} + C_3 \hat{U}_N^{(3)} + P\left(\frac{1}{N}\right),$$

where $\hat{U}_N^{(i)}$ is obtained from $U_N^{(i)}$ by centering the observations at the respective sample medians. Consider the statistic $\hat{U}_N^{(3)}$. We have,

$$A^{(3)}(t_1, t_2) = E\varphi(X - t_1, Y - t_2)$$
$$= \int F(x + t_1)\, dG(x + t_2),$$

$$W^{(3)}(x, y, t_1, t_2) = \varphi(x - t_1, y - t_2) - A^{(3)}(t_1, t_2).$$

It can then be shown that

$$E|W^{(3)}(X, Y, t_1, 0) - W^{(3)}(X, Y, 0, 0)| \leq 2at_1$$

and

$$E|W^{(3)}(X, Y, 0, t_2) - W^{(3)}(X, Y, 0, 0)| \leq 2bt_2 .$$

Condition (B_3) of Theorem 3.3 is thus satisfied. Exactly in the same manner, it can be shown that the condition (B_3) is also satisfied by the statistics $\hat{U}_N^{(1)}$ and $\hat{U}_N^{(2)}$. Condition (B_4) is also easily seen to be satisfied. Also we have

$$A^{(2)}(t_1, t_2) = E\psi(X_i - t_1, Y_j - t_2, Y_k - t_2)$$
$$= \int F(x + t_1) G(x + t_2)\, dG(x + t_2),$$

$$A^{(1)}(t_1, t_2) = E\psi(X_i - t_1, X_j - t_1, Y_k - t_2)$$
$$= \int F^2(x + t_1)\, dG(x + t_2),$$

$$\left.\frac{\partial A^{(1)}(t_1, t_2)}{\partial t_1}\right|_{t_1 = t_2 = 0} = 2 \int F(x) f(x) g(x)\, dx,$$

$$\left.\frac{\partial A^{(2)}(t_1, t_2)}{\partial t_1}\right|_{t_1 = t_2 = 0} = \int G(x) f(x) g(x)\, dx,$$

$$\left.\frac{\partial A^{(3)}(t_1, t_2)}{\partial t_1}\right|_{t_1 = t_2 = 0} = \int f(x) g(x)\, dx.$$

learly,

$$\sum_{\gamma=1}^{3} C_\gamma \frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_1}\bigg|_{t_1=t_2=0} \neq 0.$$

imilarly, it is easy to see that

$$\sum_{\gamma=1}^{3} C_\gamma \frac{\partial A^{(\gamma)}(t_1, t_2)}{\partial t_2}\bigg|_{t_1=t_2=0} \neq 0.$$

[ence, it follows as a consequence of Theorem 4.3 that the statistics $\hat{M}$ and $\hat{M}$ do not have the same asymptotic normal distribution. It follows that the est based on the statistic $\hat{M}$ is not asymptotically distribution free.

6. Small sample behavior of the proposed test. It was shown in the previous ection that the test statistic $\hat{T}$ is asymptotically distribution free. We will now ;ive some idea regarding the small sample behavior of this test by considering he simplest possible case, namely $m = n = 3$. The computations involved even n this relatively simple case are very extensive. We will consider the one sided test of the hypothesis

$$H:\delta = 1,$$

$$A:\delta > 1.$$

We will consider some special alternatives and obtain the size and the relative efficiency of the Test $\hat{T}$ with respect to the corresponding best test for each of these alternatives. These results are presented in Table 1.

TABLE 1

Population	Size of $\hat{T}$ test	Relative efficiency of $\hat{T}$ test w r t. the corresponding best test for—		
		$\delta = 2$	$\delta = 3$	$\delta = 4$
Normal	0.23	0.83	0.76	
Uniform	0.25	0.70	0.68	0 68
Double exponential....	0.25	0.92	0.81	0.81

From the above results we see that the size of the test remains more or less constant. The test is highly efficient for exponential alternatives and moderately so for normal and uniform alternatives.

7. Acknowledgment. I wish to express my deepest gratitude to Professors Erich Lehmann and Lucien LeCum, who encouraged me to work on this problem, suggested the topic and gave generous help and guidance during the course of the entire work.

REFERENCES

[1] D. A. S. FRASER, "Non-parametric theory: scale and location parameters," *Can. J. Math.*, Vol. 6 (1954), pp. 45–68.

[2] Balkrishna V. Sukhatme, "On certain two sample nonparametric tests for comparing variances," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 189–194.
[3] A. Mood, "On the asymptotic efficiency of certain nonparametric two-sample tests," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 514–522.
[4] E. S. Pearson, "The analysis of variance in cases of non-normal variation," *Biometrika*, Vol. 23 (1931), pp. 114–134.
[5] R. C. Geary, "Testing for normality," *Biometrika*, Vol. 34 (1947), pp. 209–241.
[6] A. K. Gayen, "The distribution of the variance ratio in random samples of any size drawn from non-normal universes," *Biometrika*, Vol. 37 (1950), pp. 236–255.
[7] G. E. P. Box, "Non-normality and tests on variances," *Biometrika*, Vol. 40 (1953), pp. 318–335.
[8] G. E. P. Box and S. L. Anderson, "Robust tests for variances and effect on non-normality and variance heterogeneity on standard tests," Institute of Statistics, North Carolina, Mimeographed Series No. 101 (1954).
[9] E. L. Lehmann, unpublished.
[10] W. Hoeffding, "A class of statistics with asymptotically normal distribution," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 293–325.
[11] Balkrishna V. Sukhatme, "On the joint asymptotic distribution of the median and a U-statistic," J. Roy. Stat. Soc., B, Vol. 19 (1957), pp. 144–148.
[12] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

ESTIMATION OF LOCATION AND SCALE PARAMETERS BY ORDER STATISTICS FROM SINGLY AND DOUBLY CENSORED SAMPLES[1] PART II. TABLES FOR THE NORMAL DISTRIBUTION FOR SAMPLES OF SIZE $11 \leqq n \leqq 15$

By A. E. Sarhan and B. G. Greenberg

University of North Carolina

1. Introduction. In a previous paper [2], estimation of the mean and standard deviation from singly and doubly censored samples drawn from the normal distribution were considered for samples $n \leqq 10$. The generalization of an alternative estimate for these parameters was also obtained.

In the present work, all calculations and tables obtained for the corresponding items in Part I are extended up to $n \leqq 15$.

The method is to obtain the best linear unbiased estimates of the mean and standard deviation by taking the best linear combination of the ordered observations. The variances and covariances of the order statistics for samples $11 \leqq n \leqq 15$ which are required in carrying out these calculations are obtained from Table I in [2].

Further investigation of the efficiency of the alternative estimate under varied degrees of censoring shows that the alternative estimate proposed by Gupta [1] is better than previously supposed when judged by doubly censored samples rather than singly censored samples alone.

2. Tables. Table I gives the coefficients for the best linear estimates of the mean and standard deviation for the normal population from samples of size $11 \leqq n \leqq 15$ undergoing all possible conditions of Type II censoring. Estimation from complete or singly censored samples are simply special cases and are given in the table

$$(r_1 = r_2 = 0, \text{and } r_1 \text{ or } r_2 = 0).$$

The best linear estimates of the mean and standard deviations are obtained by using

$$\mu^* = \sum_{i=r_1+1}^{n-r_2} a_{1i} y_{(i)},$$

$$\sigma^* = \sum_{i=r_1+1}^{n-r_2} a_{2i} y_{(i)},$$

where

$$y_{(1)} < y_{(2)} < \cdots, < y_{(n)}.$$

Received August 29, 1956; revised September 19, 1957.

[1] Sponsored by the Office of Ordnance Research, U. S. Army.

11

Table I. The coefficients of the most efficient linear systematic statistics of the mean and standard deviation in censored samples of sizes $11 \leq n \leq 15$ from normal populations

	$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$r_1 = 1$, r_2
μ^*	.0909	.0909	.0909	.0909	.0909	.0909	.0909	.0909	.0909	.0909	.0909	
σ^*	-.1883	-.1115	-.0760	-.0481	-.0234	.0000	.0234	.0481	.0760	.1115	.1883	
μ^*	.0781	.0841	.0869	.0891	.0910	.0928	.0945	.0963	.0982	.1891		
σ^*	-.2149	-.1256	-.0843	-.0519	-.0233	.0038	.0309	.0593	.0911	.3149		
μ^*	.0592	.0744	.0814	.0869	.0917	.0962	.1005	.1049	.3047	3.1882	-2.1882	μ^* 8
σ^*	-.2463	-.1417	-.0934	-.0555	-.0220	.0095	.0409	.0736	.4349	3.0023	-3.0023	σ^*
μ^*	.0320	.0609	.0741	.0845	.0935	.1020	.1101	.4430	1.8131	-.0776	-.7355	μ^* 7
σ^*	-.2852	-.1610	-.1038	-.0589	-.0194	.0178	.0545	.5562	1.8473	-.3251	-1.5222	σ^*
μ^*	-.0082	.0415	.0642	.0820	.0974	.1116	.6116	1.2227	.0631	.0019	-.2877	μ^* 6
σ^*	-.3357	-.1854	-.1163	-.0621	-.0146	.0299	.6842	1.3689	-.1119	-.2362	-1.0208	σ^*
μ^*	-.0698	.0128	.0504	.0797	.1049	.8220	.8646	.1026	.0738	.0407	-.0818	μ^* 5
σ^*	-.4045	-.2175	-.1317	-.0647	-.0061	.8246	1.0717	-.0195	-.0985	-.1881	-.7656	σ^*
μ^*	-.1702	-.0323	.0303	.0786	1.0937	.6157	.1116	.0971	.0813	.0632	.0311	μ^* 4
σ^*	-.5053	-.2627	-.1519	-.0657	.9857	.8517	.0301	-.0272	-.0882	-.1569	-.6095	σ^*
μ^*	-.3516	-.1104	-.0016	1.4636	.4301	.1093	.1022	.0948	.0868	.0776	.0992	μ^* 3
σ^*	-.6687	-.3331	-.1807	1.1825	.6715	.0611	.0154	-.0307	-.0796	-.1345	-.5032	σ^*
μ^*	-.7445	-.2712	2.0157	.2859	.1025	.0998	.0970	.0941	.0909	.0872	.1426	μ^* 2
σ^*	-.9862	-.4630	1.4492	.5130	.0836	.0440	.0061	-.0320	-.0722	-.1173	-.4252	σ^*
μ^*	-2.0245	3.0245	.1710	.0938	.0940	.0941	.0942	.0941	.0940	.0938	.1710	μ^* 1
σ^*	-1.9065	1.9065	.3647	.1033	.0656	.0320	.0000	-.0320	-.0656	-.1033	-.3647	σ^*

$r_1 = 2$, r_2	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$r_1 = 3$, r_2
2 μ^*	.2529	.0986	.0990	.0990	.0990	.0986	.2529	
2 σ^*	-.6113	-.0952	-.0465	.0000	.0465	.0952	.6113	
3 μ^*	.1978	.0927	.0983	.1032	.1079	.4001		
3 σ^*	-.7445	-.1095	-.0482	.0102	.0682	.8239		
4 μ^*	.1031	.0834	.0981	.1117	.6036	1.9486	-.9486	μ^* 6
4 σ^*	-.9395	-.1287	-.0484	.0276	1.0890	4.2179	-4.2179	σ^*
5 μ^*	-.0736	.0675	.0999	.9062	.9504	.0966	-.0470	μ^* 5
5 σ^*	-1.2584	-.1573	-.0453	1.4610	2.2169	-.1019	-2.1150	σ^*
6 μ^*	-.4631	.0359	1.4273	.5616	.1125	.1024	.2235	μ^* 4
6 σ^*	-1.8863	-.2083	2.0946	1.4723	.0202	-.0867	-1.4057	σ^*
7 μ^*	-1.7312	2.7312	.3413	.1057	.1059	.1057	.3413	μ^* 3
7 σ^*	-3.7473	3.7473	1.0451	.0764	.0000	-.0764	-1.0451	σ^*
			$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	

$r_1 = 4$, r_2	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$
4 μ^*	.4421	.1157	.4421
4 σ^*	-2.2233	.0000	2.2233
5 μ^*	.0000	1.0000	
5 σ^*	-4.4466	4.4466	

$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$	$r_1 = 1$, r_2
.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833	.0833	
.1748	-.1061	-.0749	-.0506	-.0294	-.0097	.0097	.0294	.0506	.0749	.1061	.1748	
.0726	.0775	.0796	.0813	.0828	.0842	.0855	.0868	.0882	.0896	.1719		
.1972	-.1185	-.0827	-.0548	-.0305	-.0079	.0142	.0367	.0608	.0881	.2919		
.0574	.0693	.0747	.0789	.0825	.0859	.0891	.0923	.0956	.2745	3.4554	-2.1554	μ^* 9
.2232	-.1324	-.0911	-.0590	-.0310	-.0050	.0203	.0461	.0733	.4020	3.0970	-3.0970	σ^*
.0360	.0581	.0682	.0759	.0827	.0888	.0948	.1006	.3950	1.9966	-.1242	-.8725	μ^* 8
.2545	-.1487	-.1007	-.0633	-.0308	-.0007	.0286	.0582	.5119	1.9252	-.3516	-1.5706	σ^*
.0057	.0428	.0595	.0724	.0836	.0938	.1036	.5336	1.3737	.0373	-.0298	-.3812	μ^* 7
.2937	-.1686	-.1119	-.0678	-.0296	.0058	.0400	.6259	1.4442	-.1345	-.2554	-1.0542	σ^*
.0382	.0210	.0477	.0684	.0861	.1022	.7128	.9961	.0870	.0542	.0162	-.1536	μ^* 6
.3448	-.1939	-.1255	-.0726	-.0267	.0155	.7479	1.1482	-.0390	-.1151	-.2024	-.7918	σ^*
.1048	-.0109	.0313	.0637	.0915	.9292	.7331	.1018	.0844	.0652	.0430	-.0275	μ^* 5
.4146	-.2274	-.1428	-.0774	-.0210	.8833	.9313	.0121	-.0423	-.1011	-.1684	-.6316	σ^*
.2125	-.0609	.0070	.0589	1.2075	.5361	.1035	.0939	.0839	.0729	.0601	.0496	μ^* 4
.5171	-.2749	-.1659	-.0820	1.0399	.7559	.0435	.0009	-.0430	-.0903	-.1442	-.5228	σ^*
.4059	-.1472	-.0321	1.5852	.3819	.0996	.0946	.0896	.0843	.0785	.0717	.0998	μ^* 3
.6836	-.3493	-.1996	1.2324	.6042	.0654	.0295	-.0059	-.0423	-.0814	-.1259	-.4435	σ^*
.8225	-.3249	2.1474	.2580	.0932	.0912	.0892	.0872	.0851	.0827	.0798	.1336	μ^* 2
1.0075	-.4874	1.4948	.4660	.0827	.0503	.0198	-.0102	-.0410	-.0738	-.1112	-.3825	σ
2.1728	3.1728	.1566	.0856	.0858	.0859	.0860	.0860	.0859	.0858	.0856	.1566	μ^* 1
1.9474	1.9474	.3334	.0989	.0671	.0391	.0128	-.0128	-.0391	-.0671	-.0989	-.3334	σ^*
		$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	$y_{(3)}$	$y_{(2)}$	

$r_1 = 2$, r_2	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$r_1 = 3$, r_2
2 μ^*	.2309	.0895	.0898	.0899	.0899	.0898	.0895	.2309	
σ^*	-.5437	-.0933	-.0545	-.0179	.0179	.0545	.0933	.5437	
3 μ^*	.1883	.0840	.0880	.0914	.0947	.0978	.3558		
σ^*	-.6440	-.1061	-.0587	-.0142	.0292	.0733	.7204		
4 μ^*	.1206	.0759	.0858	.0947	.1032	.5197	2.3903	-1.3903	μ^* 7
σ^*	-.7811	-.1225	-.0631	-.0076	.0465	.9278	4.4525	-4.4525	σ^*
5 μ^*	.0076	.0633	.0832	.1015	.7444	1.2001	.0699	-.2700	μ^* 6
σ^*	-.9829	-.1451	-.0676	.0045	1.1911	2.3763	-.1421	-2.2342	σ^*
6 μ^*	-.1993	.0417	.0806	1.0771	.7412	.1016	.0836	.0736	μ^* 5
σ^*	-1.3139	-.1796	-.0719	1.5654	1.6109	-.0111	-.1124	-1.4874	σ^*
7 μ^*	-.6489	-.0018	1.6507	.4812	.1029	.0970	.0907	.2282	μ^* 4
σ^*	-1.9673	-.2429	2.2102	1.1775	.0488	-.0223	-.0950	-1.1090	σ^*
8 μ^*	-2.0971	3.0971	.3097	.0950	.0953	.0953	.0950	.3097	μ^* 3
σ^*	-3.9063	3.9063	.8780	.0828	.0273	-.0273	-.0828	-.8780	σ^*
			$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	

$r_1 = 4$, r_2	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$r_1 = 5$, r_2
4 μ^*	.3973	.1027	.1027	.3973	
σ^*	-1.5849	-.0497	.0497	1.5849	
5 μ^*	.1962	.1034	.7004		
σ^*	-2.3874	-.0468	2.4342		
6 μ^*	-.4893	1.4893	.5000	.5000	μ^* 5
σ^*	-4.7696	4.7696	4.8738	-4.8738	σ^*
			$y_{(7)}$	$y_{(6)}$	

= 13

Table I. (Continued)

$r_1 = 0$, r_2		$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$	$y_{(13)}$		$r_1 = 1$, r_2
0	μ^*	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769	.0769		
	σ^*	-.1632	-.1013	-.0735	-.0520	-.0335	-.0164	.0000	.0164	.0335	.0520	.0735	.1013	.1632		
1	μ^*	.0679	.0718	.0735	.0749	.0761	.0771	.0781	.0792	.0802	.0813	.0824	.1576			
	σ^*	-.1824	-.1122	-.0806	-.0563	-.0353	-.0160	.0026	.0212	.0404	.0612	.0850	.2724			
2	μ^*	.0552	.0648	.0691	.0724	.0752	.0778	.0803	.0827	.0852	.0877	.2497	3.7044	-2.7044	μ^*	10
	σ^*	-.2043	-.1243	-.0884	-.0607	-.0368	-.0148	.0063	.0273	.0490	.0723	.3743	3.1823	-3.1823	σ^*	
3	μ^*	.0380	.0555	.0633	.0693	.0745	.0792	.0836	.0880	.0924	.3564	2.1686	-.1687	-.9999	μ^*	9
	σ^*	-.2301	-.1382	-.0970	-.0653	-.0379	-.0128	.0113	.0352	.0598	.4750	1.9951	-.3810	-1.6141	σ^*	
4	μ^*	.0144	.0430	.0557	.0655	.0739	.0816	.0898	.0958	.4813	1.5160	.0123	-.0600	-.4683	μ^*	8
	σ^*	-.2615	-.1549	-.1071	-.0703	-.0386	-.0095	.0192	.0456	.5781	1.5114	-.1546	-.2727	-1.0841	σ^*	
5	μ^*	-.0185	.0259	.0457	.0610	.0740	.0857	.0968	.6294	1.1209	.0712	.0352	-.0070	-.2203	μ^*	7
	σ^*	-.3011	-.1754	-.1191	-.0758	-.0386	-.0046	.0278	.6867	1.2161	-.0561	-.1298	-.2151	-.8151	σ^*	
6	μ^*	-.0659	.0020	.0322	.0553	.0750	.0928	.8085	.8453	.0916	.0716	.0496	.0239	-.0820	μ^*	6
	σ^*	-.3528	-.2015	-.1339	-.0819	-.0374	.0032	.8042	1.0015	-.0034	-.0556	-.1127	-.1786	-.6512	σ^*	
7	μ^*	-.1371	-.0330	.0132	.0484	.0784	1.0301	.6382	.0969	.0853	.0730	.0595	.0437	.0035	μ^*	5
	σ^*	-.4236	-.2363	-.1528	-.0888	-.0341	.9355	.8296	.0287	-.0116	-.0538	-.0998	-.1528	-.5401	σ^*	
8	μ^*	-.2516	-.0876	-.0151	.0400	1.3143	.4755	.0955	.0889	.0820	.0746	.0666	.0572	.0597	μ^*	4
	σ^*	-.5276	-.2859	-.1785	-.0964	1.0884	.6826	.0505	.0172	-.0164	-.0515	-.0896	-.1335	-.4594	σ^*	
9	μ^*	-.4561	-.1817	-.0610	1.6988	.3439	.0912	.0876	.0840	.0802	.0763	.0719	.0668	.0983	μ^*	3
	σ^*	-.6969	-.3638	-.2165	1.2773	.5509	.0669	.0377	.0093	-.0192	-.0489	-.0811	-.1181	-.3976	σ^*	
10	μ^*	-.8946	-.3753	2.2699	.2353	.0853	.0839	.0824	.0809	.0794	.0778	.0759	.0737	.1254	μ^*	2
	σ^*	-1.0266	-.5094	1.5360	.4278	.0808	.0536	.0283	.0039	-.0207	-.0462	-.0738	-.1055	-.3483	σ^*	
11	μ^*	-2.3101	3.3101	.1445	.0788	.0790	.0791	.0791	.0791	.0791	.0791	.0790	.0788	.1445	μ^*	1
	σ^*	-1.9845	1.9845	.3076	.0947	.0673	.0434	.0213	.0000	-.0213	-.0434	-.0673	-.0947	-.3076	σ^*	
				$y_{(12)}$	$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	$y_{(3)}$	$y_{(2)}$		

$n = 13$ Table I. (Continued)

$r_1 = 2$ r_2		$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$		$r_1 = 3$ r_2
2	μ^*	.2125	.0819	.0822	.0823	.0823	.0823	.0822	.0819	.2125		
	σ^*	-.4911	-.0904	-.0584	-.0287	.0000	.0287	.0584	.0904	.4911		
3	μ^*	.1784	.0770	.0799	.0825	.0848	.0871	.0892	.3211			
	σ^*	-.5698	-.1017	-.0636	-.0283	.0058	.0398	.0748	.6431			
4	μ^*	.1272	.0700	.0770	.0833	.0892	.0949	.4584	2.8102	-1.8102	μ^*	8
	σ^*	-.6724	-.1157	-.0695	-.0266	.0146	.0555	.8142	4.6615	-4.6615	σ^*	
5	μ^*	.0479	.0597	.0732	.0854	.0969	.6370	1.4405	.0415	-.4820	μ^*	7
	σ^*	-.8133	-.1341	-.0763	-.0229	.0282	1.0184	2.5175	-.1774	-2.3402	σ^*	
6	μ^*	-.0822	.0437	.0680	.0901	.8805	.9169	.0883	.0639	-.0691	μ^*	6
	σ^*	-1.0214	-.1597	-.0846	-.0154	1.2810	1.7328	-.0380	-.1350	-1.5598	σ^*	
7	μ^*	-.3174	.0164	.0607	1.2404	.6209	.0970	.0865	.0752	.1204	μ^*	5
	σ^*	-1.3635	-.1994	-.0952	1.6581	1.2929	.0255	-.0417	-.1115	-1.1652	σ^*	
8	μ^*	-.8233	-.0391	1.8624	.4246	.0939	.0902	.0862	.0820	.2231	μ^*	4
	σ^*	-2.0397	-.2735	2.3132	.9927	.0610	.0095	-.0423	-.0958	-.9251	σ^*	
9	μ^*	-2.4408	3.4408	.2838	.0863	.0866	.0866	.0866	.0863	.2838	μ^*	3
	σ^*	-4.0488	4.0488	.7622	.0841	.0414	.0000	-.0414	-.0841	-.7622	σ^*	
				$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$		

$r_1 = 4$ r_2		$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$		$r_1 = 5$ r_2
4	μ^*	.3615	.0923	.0924	.0923	.3615		
	σ^*	-1.2545	-.0674	.0000	.0674	1.2545		
5	μ^*	.2377	.0900	.0977	.5745			
	σ^*	-1.6839	-.0755	.0184	1.7410			
6	μ^*	-.0422	.0861	.9562	· · · ·	.0000	μ^*	6
	σ^*	-2.5323	-.0873	2.6196	5.2487	-5.2487	σ^*	
7	μ^*	-.9632	1.9632	.4498	.1003	.4498	μ^*	5
	σ^*	-5.0555	5.0555	2.6243	.0000	-2.6243	σ^*	
				$y_{(8)}$	$y_{(7)}$	$y_{(6)}$		

n = 14

Table I. (Continued)

$r_1 = 0$ r_2		$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$	$y_{(13)}$	$y_{(14)}$	
0	μ^*	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	.0714	
	σ^*	-.1532	-.0968	-.0717	-.0526	-.0362	-.0212	-.0070	.0070	.0212	.0362	.0526	.0717	.0968	.1532	
1	μ^*	.0637	.0669	.0683	.0694	.0704	.0712	.0721	.0728	.0736	.0745	.0753	.0762	.1155		
	σ^*	-.1698	-.1065	-.0784	-.0568	-.0384	-.0216	-.0056	.0100	.0259	.0426	.0609	.0820	.2556		
2	μ^*	.0530	.0609	.0643	.0670	.0692	.0713	.0732	.0751	.0770	.0789	.0809	.2291	3.9374	-2.9374	μ^*
	σ^*	-.1885	-.1171	-.0854	-.0612	-.0404	-.0215	-.0036	.0140	.0319	.0505	.0707	.3506	3.2597	-3.2597	σ^*
3	μ^*	.0388	.0529	.0592	.0639	.0680	.0717	.0752	.0785	.0819	.0852	.3247	2.3304	-.2112	-1.1191	μ^*
	σ^*	-.2102	-.1292	-.0933	-.0658	-.0423	-.0209	-.0006	.0192	.0393	.0601	.4438	2.0584	-.4048	-1.6536	σ^*
4	μ^*	.0199	.0426	.0526	.0602	.0667	.0726	.0782	.0835	.0887	.4350	1.6505	-.0121	-.0838	-.5496	μ^*
	σ^*	-.2361	-.1434	-.1023	-.0709	-.0440	-.0195	.0035	.0260	.0487	.5382	1.5721	-.1726	-.2832	-1.1112	σ^*
5	μ^*	-.0057	.0288	.0440	.0557	.0655	.0744	.0828	.0903	.5637	1.2394	.0556	.0167	-.0291	-.2827	μ^*
	σ^*	-.2678	-.1604	-.1129	-.0765	-.0455	-.0174	.0092	.0350	.6363	1.2771	-.0713	-.1430	-.2267	-.8362	σ^*
6	μ^*	-.0411	.0102	.0328	.0500	.0646	.0777	.0899	.7159	.9525	.0811	.0591	.0345	.0057	-.1329	μ^*
	σ^*	-.3077	-.1815	-.1256	-.0829	-.0466	-.0137	.0172	.7407	1.0642	-.0171	-.0675	-.1231	-.1878	-.6658	σ^*
7	μ^*	-.0915	-.0158	.0175	.0429	.0643	.0835	.8992	.7365	.0897	.0765	.0623	.0466	.0281	-.0397	μ^*
	σ^*	-.3599	-.2084	-.1414	-.0903	-.0469	-.0077	.8546	.8949	.0158	-.0227	-.0635	-.1084	-.1606	-.5555	σ^*
8	μ^*	-.1670	-.0537	-.0040	.0338	.0655	1.1255	.5661	.0903	.0827	.0743	.0652	.0552	.0435	.0222	μ^*
	σ^*	-.4317	-.2444	-.1618	-.0990	-.0457	.9825	.7515	.0379	.0065	-.0256	-.0596	-.0970	-.1402	-.4734	σ^*
9	μ^*	-.2879	-.1127	-.0360	.0218	1.4448	.4277	.0883	.0834	.0784	.0733	.0677	.0616	.0544	.0651	μ^*
	σ^*	-.5372	-.2959	-.1898	-.1094	1.1322	.6244	.0540	.0270	.0003	-.0271	-.0560	-.0876	-.1242	-.4107	σ^*
10	μ^*	-.5027	-.2142	-.0886	1.8054	.3129	.0840	.0812	.0785	.0758	.0729	.0699	.0665	.0625	.0957	μ^*
	σ^*	-.7091	-.3771	-.2318	1.3180	.5075	.0669	.0426	.0191	-.0040	-.0276	-.0525	-.0797	-.1112	-.3610	σ^*
11	μ^*	-.9616	-.4228	2.3843	.2163	.0787	.0776	.0765	.0753	.0742	.0730	.0717	.0703	.0685	.1180	μ^*
	σ^*	-1.0441	-.5293	1.5734	.3961	.0785	.0552	.0338	.0133	-.0070	-.0275	-.0492	-.0728	-.1002	-.3203	σ^*
12	μ^*	-2.4378	3.4378	.1341	.0730	.0731	.0732	.0733	.0733	.0733	.0733	.0732	.0731	.0730	.1341	μ^*
	σ^*	-2.0182	2.0182	.2858	.0907	.0667	.0459	.0270	.0089	-.0089	-.0270	-.0459	-.0667	-.0907	-.2858	σ^*
				$y_{(13)}$	$y_{(12)}$	$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	$y_{(3)}$	$y_{(2)}$	

n = 14

Table I. (Continued)

$r_1 = 2$, r_2		$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$		$r_1 = 3$, r_2
2	μ^*	.1969	.0755	.0758	.0759	.0759	.0759	.0759	.0758	.0755	.1969		
	σ^*	-.4488	-.0872	-.0602	-.0354	-.0117	.0117	.0354	.0602	.0872	.4488		
3	μ^*	.1689	.0711	.0734	.0753	.0771	.0788	.0804	.0820	.2930			
	σ^*	-.5125	-.0972	-.0656	-.0367	-.0090	.0182	.0457	.0744	.5827			
4	μ^*	.1287	.0651	.0703	.0749	.0792	.0834	.0874	.4111	3.2094	-2.2094	μ^*	9
	σ^*	-.5926	-.1092	-.0718	-.0375	-.0048	.0273	.0596	.7291	4.8497	-4.8497	σ^*	
5	μ^*	.0695	.0565	.0661	.0748	.0829	.0907	.5593	1.6712	.0123	-.6835	μ^*	8
	σ^*	-.6975	-.1244	-.0791	-.0376	.0018	.0403	.8965	2.6442	-.2087	-2.4355	σ^*	
6	μ^*	-.0206	.0440	.0606	.0755	.0895	.7510	1.0874	.0734	.0439	-.2047	μ^*	7
	σ^*	-.8420	-.1444	-.0881	-.0365	.0123	1.0987	1.8416	-.0617	-.1552	-1.6247	σ^*	
7	μ^*	-.1668	.0247	.0527	.0779	1.0115	.7583	.0893	.0750	.0596	.0178	μ^*	6
	σ^*	-1.0559	-.1728	-.0996	-.0330	1.3612	1.3951	.0053	-.0588	-.1262	-1.2153	σ^*	
8	μ^*	-.4285	-.0084	.0405	1.3964	.5397	.0905	.0837	.0766	.0690	.1405	μ^*	5
	σ^*	-1.4081	-.2172	-.1160	1.7413	1.0932	.0424	-.0061	-.0555	-.1074	-.9667	σ^*	
9	μ^*	-.9875	-.0756	2.0631	.3818	.0860	.0835	.0809	.0781	.0750	.2147	μ^*	4
	σ^*	-2.1051	-.3010	2.4061	.8643	.0660	.0266	-.0125	-.0522	-.0937	-.7987	σ^*	
10	μ^*	-2.7647	3.7647	.2621	.0791	.0794	.0795	.0795	.0794	.0791	.2621	μ^*	3
	σ^*	-4.1778	4.1778	.6765	.0831	.0489	.0162	-.0162	-.0489	-.0831	-.6765	σ^*	
		$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$				

$r_1 = 4$, r_2		$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$		$r_1 = 5$, r_2
4	μ^*	.3320	.0839	.0841	.0841	.0839	.3320		
	σ^*	-1.0495	-.0738	-.0244	.0244	.0738	1.0495		
5	μ^*	.2456	.0807	.0857	.0903	.4978			
	σ^*	-1.3227	-.0839	-.0195	.0439	1.3821			
6	μ^*	.0837	.0755	.0895	.7512	1.4921	-.4921	*	7
	σ^*	-1.7721	-.0983	-.0088	1.8791	5.5823	-5.5823	σ^*	
7	μ^*	-.2720	.0657	1.2063	.7061	.0910	.2029	*	6
	σ^*	-2.6618	-.1229	2.7848	2.8338	-.0406	-2.7931	σ^*	
8	μ^*	-1.4198	2.4198	.4098	.0902	.0902	.4098	*	5
	σ^*	-5.3115	5.3115	1.8560	.0441	-.0441	-1.8560	σ^*	
				$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$		

$r_1 = 6$, r_2		$y_{(7)}$	$y_{(8)}$
6	μ^*	.5000	.5000
	σ^*	-5.6716	5.6716

Table I. (Continued)

n = 15

$r_1 = 0$, r_2		$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$	$y_{(13)}$	$y_{(14)}$	$y_{(15)}$
0	μ^*	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667	.0667
	σ^*	-.1444	-.0927	-.0699	-.0526	-.0379	-.0247	-.0122	.0000	.0122	.0247	.0379	.0526	.0699	.0927	.1444
1	μ^*	.0599	.0627	.0639	.0648	.0655	.0662	.0669	.0675	.0682	.0688	.0695	.0702	.0709	.1351	
	σ^*	-.1590	-.1013	-.0760	-.0568	-.0404	-.0256	-.0116	.0019	.0154	.0293	.0440	.0602	.0791	.2409	
2	μ^*	.0508	.0574	.0602	.0624	.0642	.0659	.0675	.0690	.0704	.0719	.0735	.0751	.2116	4.1564	-3.1564
	σ^*	-.1752	-.1108	-.0825	-.0610	-.0427	-.0262	-.0106	.0044	.0195	.0349	.0512	.0690	.3300	3.3306	-3.3306
3	μ^*	.0390	.0506	.0556	.0595	.0628	.0657	.0685	.0711	.0737	.0763	.0790	.2982	2.4829	-.2519	-1.2311
	σ^*	-.1937	-.1214	-.0897	-.0655	-.0450	-.0265	-.0091	.0078	.0246	.0417	.0598	.4169	2.1162	-.4265	-1.6897
4	μ^*	.0234	.0418	.0498	.0560	.0611	.0658	.0701	.0743	.0784	.0824	.3969	1.7779	-.0356	-.1163	-.6260
	σ^*	-.2154	-.1336	-.0977	-.0705	-.0473	-.0264	-.0068	.0122	.0310	.0502	.5042	1.6273	-.1890	-.3024	-1.1359
5	μ^*	.0030	.0305	.0425	.0516	.0593	.0663	.0727	.0789	.0849	.5104	1.3522	.0403	-.0011	-.0501	-.3413
	σ^*	-.2414	-.1481	-.1071	-.0760	-.0496	-.0258	-.0035	.0180	.0393	.5940	1.3324	-.0850	-.1549	-.2373	-.8553
6	μ^*	-.0244	.0155	.0330	.0462	.0574	.0674	.0767	.0856	.6425	1.0549	.0706	.0467	.0200	-.0115	-.1807
	σ^*	-.2733	-.1654	-.1181	-.0822	-.0518	-.0244	.0012	.0258	.6882	1.1208	-.0293	-.0781	-.1325	-.1963	-.6347
7	μ^*	-.0621	-.0046	.0205	.0395	.0555	.0698	.0830	.7983	.8308	.0823	.0677	.0519	.0342	.0134	-.0803
	σ^*	-.3136	-.1870	-.1315	-.0894	-.0538	-.0219	.0079	.7892	.9536	.0045	-.0326	-.0722	-.1162	-.1677	-.5694
8	μ^*	-.1155	-.0326	.0036	.0309	.0539	.0743	.9854	.6537	.0857	.0764	.0666	.0561	.0443	.0305	-.0131
	σ^*	-.3664	-.2146	-.1482	-.0979	-.0555	-.0174	.9001	.8129	.0269	-.0029	-.0339	-.0670	-.1036	-.1464	-.1859
9	μ^*	-.1950	-.0732	-.0203	.0196	.0531	1.2157	.5093	.0848	.0789	.0728	.0664	.0595	.0518	.0428	.0338
	σ^*	-.4390	-.2518	-.1700	-.1082	-.0562	1.0252	.6892	.0430	.0177	-.0078	-.0342	-.0624	-.0935	-.1298	-.4223
10	μ^*	-.3217	-.1364	-.0560	.0043	1.5097	.3890	.0818	.0781	.0744	.0706	.0666	.0623	.0575	.0519	.0677
	σ^*	-.5459	-.3050	-.2002	-.1211	1.1722	.5767	.0555	.0331	.0111	-.0110	-.0339	-.0582	-.0850	-.1163	-.3721
11	μ^*	-.5462	-.2448	-.1148	1.9058	.2873	[illegible]	.0756	.0735	.0714	.0693	.0671	.0647	.0620	.0587	.0927
	σ^*	-.7201	-.3892	-.2458	1.3552	.471	[illegible]	53	.0256	.0062	-[illegible]	-.0331	-.0543	-.0778	-.1050	-.3311
12	μ^*	-1.0242	-.4676	2.4918	.2002	.073			.0704	.0695		6	.0666	.0654	.0640	.1113
	σ^*	-1.0601	-.5477	1.6077	.3693	.076			0197	.0026	-		-.0508	-.0714	-.0954	-.2968
13	μ^*	-2.5574	3.5574	.1252	.0679	.0681			2	.0682	.		82	.0681	.0679	.1252
	σ^*	-2.0493	2.0493	.2672	.0869	.0656				.0000	-.0		4	-.0656	-.0869	-.2672
				$y_{(14)}$	$y_{(13)}$	$y_{(12)}$				$y_{(8)}$	$y_{(}$			$y_{(4)}$	$y_{(3)}$	$y_{(2)}$

n = 15

$r_1 = 2$, r_2		$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$y_{(12)}$	$y_{(13)}$	$r_1 = 3$, r_2
2	μ^*	.1835	.0701	.0703	.0704	.0705	.0705	.0705	.0704	.0703	.0701	.1835	
	σ^*	-.4139	-.0838	-.0607	-.0395	-.0195	.0000	.0195	.0395	.0607	.0838	.4139	
3	μ^*	.1601	.0662	.0679	.0695	.0708	.0721	.0734	.0746	.0758	.2697		
	σ^*	-.4668	-.0927	-.0660	-.0417	-.0187	.0037	.0261	.0490	.0730	.5341		
4	μ^*	.1274	.0609	.0649	.0684	.0717	.0748	.0778	.0808	.3734	3.5892	-2.5892	μ^* 10
	σ^*	-.5314	-.1032	-.0721	-.0438	-.0170	.0090	.0349	.0612	.6624	5.0207	-5.0207	σ^*
5	μ^*	.0815	.0536	.0608	.0673	.0733	.0790	.0845	.5000	1.8923	-.0171	-.8752	μ^* 9
	σ^*	-.6129	-.1160	-.0792	-.0457	-.0142	.0164	.0468	.8049	2.7587	-.2368	-2.5219	σ^*
6	μ^*	.0150	.0436	.0555	.0662	.0762	.0857	.6579	1.2523	.0577	0238	-.3338	μ^* 8
	σ^*	-.7200	-.1323	-.0878	-.0475	-.0095	.0272	.9699	1.9396	-.0828	-.1734	-1.6834	σ^*
7	μ^*	-.0854	.0289	.0481	.0654	.0815	.8614	.8925	.0804	.0629	.0441	-.0799	μ^* 7
	σ^*	-.8679	-.1539	-.0987	-.0486	-.0017	1.1708	1.4867	-.0125	-.0741	-.1395	-1.2605	σ^*
8	μ^*	-.2467	.0063	.0375	.0655	1.1375	.6535	.0857	.0764	.0666	.0561	.0617	μ^* 6
	σ^*	-1.0872	-.1847	-.1132	-.0486	1.4336	1.1828	.0264	-.0198	-.0674	-.1178	-1.0042	σ^*
9	μ^*	-.5335	-.0324	.0205	1.5454	.4801	.0840	.0793	.0745	.0694	.0639	.1487	μ^* 5
	σ^*	-1.4486	-.2334	-.1346	1.8166	.9544	.0508	.0137	-.0235	-.0619	-.1023	-.8312	σ^*
10	μ^*	-1.1426	-.1111	2.2537	.3478	.0792	.0775	.0756	.0737	.0716	.0693	.2054	μ^* 4
	σ^*	-2.1646	-.3259	2.4905	.7692	.0678	.0363	.0055	-.0254	-.0571	-.0905	-.7058	σ^*
11	μ^*	-3.0706	4.0706	.2436	.0731	.0733	.0734	.0734	.0734	.0733	.0731	.2436	μ^* 3
	σ^*	-4.2953	4.2953	.6102	.0810	.0528	.0261	.0000	-.0261	-.0528	-.0810	-.6102	σ^*
				$y_{(12)}$	$y_{(11)}$	$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	$y_{(5)}$	$y_{(4)}$	

$r_1 = 4$, r_2		$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(10)}$	$y_{(11)}$	$r_1 = 5$, r_2
4	μ^*	.3073	.0770	.0772	.0772	.0772	.0770	.3073	
	σ^*	-.9085	-.0756	-.0374	.0000	.0374	.0756	.9085	
5	μ^*	.2424	.0736	.0770	.0802	.0832	.4435		
	σ^*	-1.1001	-.0854	-.0377	.0088	.0552	1.1592		
6	μ^*	.1339	.0686	.0774	.0859	.6342	1.9724	-.9724	μ^* 8
	σ^*	-1.3838	-.0987	-.0367	.0236	1.4956	5.8824	-5.8824	σ^*
7	μ^*	-.0644	.0601	.0791	.9252	.9607	.0776	-.0383	μ^* 7
	σ^*	-1.8515	-.1186	-.0327	2.0028	3.0212	-.0764	-2.9448	σ^*
8	μ^*	-.4928	.0435	1.1494	.5846	.0865	.0804	.2486	μ^* 6
	σ^*	-2.7788	-.1547	2.9335	2.0090	.0169	-.0670	-1.9590	σ^*
9	μ^*	-1.8586	2.8586	.3769	.0820	.0821	.0820	.3769	μ^* 5
	σ^*	-5.5431	5.5431	1.1614	.0604	.0000	-.0604	-1.4614	σ^*
				$y_{(10)}$	$y_{(9)}$	$y_{(8)}$	$y_{(7)}$	$y_{(6)}$	

$r_1 = 6$, r_2		$y_{(7)}$	$y_{(8)}$	$y_{(9)}$
6	μ^*	.4557	.0887	.4557
	σ^*	-3.0248	.0000	3.0248
7	μ^*	.0000	1.0000	
	σ^*	-6.0497	6.0497	

Table II. Variances and Covariances of the Estimates of the

of Sizes $11 \le n \le 15$

n	r_1 \ r_2		0	1	2	3	4	5	6	7	8	9	10
11	0	$V(\mu^*)$	.0909	.0929	.0966	.1031	.1143	.1342	.1718	.2504	.4493	1.2179	
		$V(\sigma^*)$	.0517	.0601	.0704	.0836	.1013	.1262	.1640	.2279	.3577	.7550	
		$Cov(\mu^*, \sigma^*)$	.0000	.0041	.0102	.0195	.0336	.0559	.0936	.1644	.3251	.8777	
	1	$V(\mu^*)$		.0943	.0975	.1033	.1143	.1356	.1805	.2929	.7093		
		$V(\sigma^*)$		.0712	.0854	.1043	.1311	.1717	.2408	.3817	.8140		
		$Cov(\mu^*, \sigma^*)$		.0000	.0066	.0172	.0343	.0637	.1194	.2453	.6696		
	2	$V(\mu^*)$			.0997	.1044	.1144	.1364	.1926	.3997			
		$V(\sigma^*)$			.1052	.1330	.1753	.2469	.3928	.8389			
		$Cov(\mu^*, \sigma^*)$			.0000	.0115	.0321	.0718	.1623	.4662			
	3	$V(\mu^*)$				.1075	.1153	.1369	.2202				
		$V(\sigma^*)$				.1764	.2496	.3982	.8512				
		$Cov(\mu^*, \sigma^*)$				.0000	.0239	.0806	.2748				
	4	$V(\mu^*)$					.1191	.1372					
		$V(\sigma^*)$					.3999	.8564					
		$Cov(\mu^*, \sigma^*)$					.0000	.0908					
12	0	$V(\mu^*)$	.0833	.0849	.0878	.0926	.1004	.1136	.1363	.1784	.2650	.4809	1.3044
		$V(\sigma^*)$	.0469	.0538	.0620	.0723	.0855	.1033	.1283	.1663	.2305	.3610	.7601
		$Cov(\mu^*, \sigma^*)$	.0000	.0033	.0081	.0152	.0254	.0406	.0645	.1045	.1791	.3469	.9202
	1	$V(\mu^*)$		.0861	.0885	.0929	.1004	.1139	.1393	.1915	.3195	.7864	
		$V(\sigma^*)$		.0627	.0737	.0878	.1067	.1336	.1745	.2438	.3854	.8195	
		$Cov(\mu^*, \sigma^*)$		.0000	.0052	.0130	.0250	.0440	.0762	.1364	.2710	.7212	
	2	$V(\mu^*)$			.0903	.0939	.1007	.1140	.1421	.2115	.4614		
		$V(\sigma^*)$			.0884	.1082	.1361	.1786	.2505	.3971	.8450		
		$Cov(\mu^*, \sigma^*)$			.0000	.0084	.0222	.0460	.0909	.1918	.5263		
	3	$V(\mu^*)$				.0963	.1018	.1140	.1457	.2617			
		$V(\sigma^*)$				.1369	.1804	.2539	.4032	.8581			
		$Cov(\mu^*, \sigma^*)$				.0000	.0153	.0454	.1141	.3438			
	4	$V(\mu^*)$					.1049	.1144	.1524				
		$V(\sigma^*)$					.2549	.4059	.8647				
		$Cov(\mu^*, \sigma^*)$					.0000	.0379	.1699				
	5	$V(\mu^*)$						.1175					
		$V(\sigma^*)$						.8667					
		$Cov(\mu^*, \sigma^*)$						.0000					

μ*) and Standard Deviation (σ*) for Censored Samples

ormal Populations

r_1 \ r_2		0	1	2	3	4	5	6	7	8	9	10	11	
0	V(μ*)	.0769	.0782	.0805	.0841	.0899	.0991	.1141	.1395	.1858	.2799	.5118	1.3868	
	V(σ*)	.0429	.0486	.0554	.0636	.0739	.0872	.1050	.1302	.1684	.2329	.3640	.7646	
	Cov(μ*, σ*)	.0000	.0027	.0066	.0121	.0198	.0309	.0472	.0725	.1146	.1925	.3668	.9589	
1	V(μ*)		.0792	.0811	.0845	.0900	.0992	.1152	.1444	.2036	.3466	.8613		
	V(σ*)		.0559	.0647	.0756	.0898	.1088	.1358	.1768	.2465	.3886	.8243		
	Cov(μ*, σ*)		.0000	.0041	.0101	.0190	.0322	.0530	.0876	.1518	.2944	.7679		
2	V(μ*)			.0826	.0854	.0903	.0992	.1159	.1500	.2324	.5237			
	V(σ*)			.0761	.0909	.1106	.1387	.1813	.2536	.4007	.8502			
	Cov(μ*, σ*)			.0000	.0064	.0162	.0320	.0587	.1084	.2184	.5803			
3	V(μ*)				.0874	.0914	.0993	.1164	.1584	.3078				
	V(σ*)				.1112	.1399	.1836	.2574	.4073	.8639				
	Cov(μ*, σ*)				.0000	.0107	.0294	.0648	.1442	.4054				
4	V(μ*)					.0939	.1002	.1167	.1780					
	V(σ*)					.1843	.2591	.4108	.8714					
	Cov(μ*, σ*)					.0000	.0216	.0716	.2397					
5	V(μ*)						.1032	.1168						
	V(σ*)						.4118	.8747						
	Cov(μ*, σ*)						.0000	.0793						
0	V(μ*)	.0714	.0725	.0743	.0772	.0816	.0883	.0988	.1155	.1435	.1938	.2950	.5419	1.[illegible]
	V(σ*)	.0395	.0444	.0500	.0568	.0650	.0753	.0886	.1065	.1318	.1702	.2350	.3666	.[illegible]
	Cov(μ*, σ*)	.0000	.0023	.0055	.0099	.0159	.0242	.0360	.0533	.0799	.1239	.2048	.3851	.[illegible]
1	V(μ*)		.0733	.0749	.0775	.0817	.0883	.0992	.1175	.1505	.2164	.3736	.9337	
	V(σ*)		.0505	.0576	.0664	.0773	.0915	.1106	.1376	.1789	.2488	.3914	.8284	
	Cov(μ*, σ*)		.0000	.0034	.0081	.0149	.0246	.0390	.0612	.0981	.1660	.3158	.8105	
2	V(μ*)			.0762	.0784	.0821	.0884	.0993	.1193	.1594	.2543	.5858		
	V(σ*)			.0668	.0781	.0929	.1127	.1408	.1836	.2561	.4038	.8547		
	Cov(μ*, σ*)			.0000	.0050	.0124	.0236	.0411	.0704	.1243	.2427	.6293		
3	V(μ*)				.0800	.0831	.0887	.0993	.1212	.1737	.3567			
	V(σ*)				.0933	.1137	.1425	.1863	.2604	.4108	.8689			
	Cov(μ*, σ*)				.0000	.0079	.0206	.0422	.0825	.1714	.4609			
4	V(μ*)					.0852	.0896	.0994	.1237	.2104				
	V(σ*)					.1430	.1875	.2626	.4148	.8770				
	Cov(μ*, σ*)					.0000	.0141	.0411	.1020	.3022				
5	V(μ*)						.0922	.0997	.1286					
	V(σ*)						.2633	.4166	.8812					
	Cov(μ*, σ*)						.0000	.0338	.1496					
6	V(μ*)							.1022						
	V(σ*)							.8825						
	Cov(μ*, σ*)							.0000						

Table 11. (Continued)

n	r_1	r_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13
15	0	$V(\mu^*)$	.0667	.0676	.0691	.0714	.0748	.0799	.0875	.0992	.1176	.1480	.2022	.3101	.5713	1.5404
		$V(\sigma^*)$	.0366	.0408	.0456	.0512	.0580	.0662	.0765	.0899	.1078	.1332	.1718	.2368	.3689	.7723
		$Cov(\mu^*,\sigma^*)$	.0000	.0019	.0046	.0082	.0130	.0195	.0284	.0409	.0590	.0868	.1325	.2163	.4020	1.0273
	1	$V(\mu^*)$		.0683	.0696	.0717	.0749	.0799	.0877	.1001	.1207	.1574	.2296	.4004	1.0037	
		$V(\sigma^*)$		.0460	.0519	.0590	.0678	.0787	.0929	.1121	.1393	.1807	.2508	.3939	.8322	
		$Cov(\mu^*,\sigma^*)$		.0000	.0028	.0067	.0120	.0194	.0299	.0453	.0689	.1079	.1791	.3355	.8496	
	2	$V(\mu^*)$			.0706	.0724	.0753	.0800	.0877	.1007	.1239	.1697	.2769	.6471		
		$V(\sigma^*)$			.0594	.0684	.0798	.0946	.1145	.1427	.1856	.2584	.4065	.8586		
		$Cov(\mu^*,\sigma^*)$			.0000	.0040	.0098	.0181	.0304	.0496	.0812	.1389	.2649	.6740		
	3	$V(\mu^*)$				.0738	.0762	.0804	.0877	.1011	.1279	.1908	.4069			
		$V(\sigma^*)$				.0802	.0953	.1157	.1446	.1886	.2629	.4138	.8731			
		$Cov(\mu^*,\sigma^*)$				.0000	.0061	.0153	.0298	.0541	.0987	.1962	.5112			
	4	$V(\mu^*)$					.0780	.0813	.0879	.1014	.1340	.2472				
		$V(\sigma^*)$					.1161	.1455	.1901	.2655	.4182	.8817				
		$Cov(\mu^*,\sigma^*)$					.0000	.0099	.0270	.0590	.1295	.3585				
	5	$V(\mu^*)$						.0835	.0886	.1016	.1487					
		$V(\sigma^*)$						.1906	.2667	.4205	.8866					
		$Cov(\mu^*,\sigma^*)$						.0000	.0196	.0644	.2126					
	6	$V(\mu^*)$							.0911	.1017						
		$V(\sigma^*)$							.4212	.8888						
		$Cov(\mu^*,\sigma^*)$							[illegible]	[illegible]						

This table is a continuation of Table II in [2]. The entries in Table I, as well as in Table II of this present paper, have been rounded to four decimal places for convenience. Readers who desire more precision may obtain copies of the original tables containing eight decimal places from the authors. The results in the eight-decimal-place table are exact to seven places but rounding may cause some of the figures in the eighth place to be a few units in error.

If the coefficients of an estimate are sought for a value of r_1 *not given in the table, the same procedure can be followed as that mentioned in Part I of this* series.

The variances of the estimates and their covariances are given in Table II in terms of σ^2. This table is a continuation of Table III in [2] and the results are given to only four decimal places for convenience

Table III shows the efficiency of the estimates for every case of censoring relative to the corresponding estimate obtained by complete samples.

3. Alternative estimate. The alternative estimate was proposed by Gupta [1] to replace the best linear estimate when sample sizes are greater than 10 and censoring was from one side only. This estimate was generalized to the case of double censoring in Part I of this present series. The variance of the alternative estimates and their efficiencies relative to the best linear estimate for samples of sizes 12 and 15 under every case of censoring are given in Table IV.

The authors know of no instance where the alternative estimates have been compared previously for sample sizes this large.

4. Comments. The conclusions mentioned in [2], Section 5, hold true here and, in fact, appear much stronger for increasing sample size. Several points are worth emphasis:

(1) *In estimation of the mean, the relative efficiency holds up—about 65 per cent or better—as long as the median value remains known. (For an even n, it is* about 70 per cent or better as long as the two middle values are uncensored.) This result was anticipated because the asymptotic efficiency of the median is $2/\pi = 63.7\%$.

Another way of presenting this same finding can be seen clearly from Fig. 1 which shows the relative efficiency of the best linear estimate of μ under all conditions of censoring a sample of size 15 from the normal distribution. Each one of the curves shows the efficiency of the estimate of the mean for a certain number of known elements $[k = n - (r_1 + r_2)]$ for all possible values of r_1 and r_2. The efficiency attains its maximum whenever the middle element is known.

(2) From Table III, one can see that, for fixed values of censoring from one side, the efficiency of the estimate of the standard deviation decreases approximately in equal amounts with each increment in the number of censored elements on the opposite side.

This is illustrated by Fig. 2 which shows the relative efficiency of the best linear estimate of σ under all conditions of censoring a sample of size 15 from the

Table III

Percentage efficiencies of the estimate of the mean (μ^*) and standard deviation (σ^*) for censored samples relative to uncensored samples in a normal population for $11 \le n \le 15$ (Continuation of Table IV in reference [2].)

n	r_1 \ r_2		0	1	2	3	4	5	6	7	8	9	10
11	0	μ^*	100.0	97.89	94.13	88.20	79.55	67.74	52.92	36.31	20.23	7.46	
		σ^*	100.0	86.02	73.43	61.82	51.04	40.95	31.51	22.68	14.45	6.85	
	1	μ^*		96.36	93.29	87.98	79.54	67.06	50.37	31.04	12.82		
		σ^*		72.55	60.53	49.55	39.44	30.10	21.47	13.54	6.35		
	2	μ^*			91.20	87.04	79.46	66.64	47.21	22.75			
		σ^*			49.14	38.86	29.48	20.93	13.16	6.16			
	3	μ^*				84.56	78.84	66.39	41.29				
		σ^*				29.31	20.71	12.98	6.07				
	4	μ^*					76.32	66.28					
		σ^*					12.93	6.04					

12	0	μ^*	100.0	98.16	94.95	90.04	82.98	73.38	61.13	46.72	31.45	17.33	6.39
		σ^*	100.0	87.18	75.57	64.84	54.80	45.39	36.53	28.18	20.33	12.98	6.17
	1	μ^*		96.78	94.13	89.74	82.97	73.14	59.83	43.52	26.08	10.60	
		σ^*		74.78	63.63	53.40	43.92	35.09	26.87	19.22	12.16	5.72	
	2	μ^*			92.25	88.75	82.75	73.08	58.63	39.40	18.06		
		σ^*			53.01	43.33	34.44	26.25	18.71	11.81	5.55		
	3	μ^*				86.49	81.89	73.08	57.21	31.85			
		σ^*				34.25	25.99	18.47	11.63	5.46			
	4	μ^*					79.43	72.84	54.68				
		σ^*					18.39	11.55	5.42				
	5	μ^*						70.91					
		σ^*						5.41					

Table III (continued)

n	r_1	r_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13
13	0	μ^*	100.0	98.36	95.60	91.44	85.55	77.61	67.43	55.13	41.39	27.48	15.03	5.55		
		σ^*	100.0	88.16	77.39	67.39	58.02	49.19	40.84	32.94	25.47	18.41	11.78	5.61		
	1	μ^*		97.11	94.80	91.08	85.51	77.55	66.80	53.27	37.78	22.19	8.93			
		σ^*		76.66	66.28	56.70	47.77	39.42	31.59	24.25	17.40	11.03	5.20			
	2	μ^*			93.10	90.09	85.17	77.55	66.38	51.28	33.10	14.69				
		σ^*			56.32	47.20	38.76	30.93	23.65	16.91	10.70	5.04				
	3	μ^*				88.03	84.20	77.43	66.10	48.58	24.99					
		σ^*				38.56	30.65	23.36	16.66	10.53	4.96					
	4	μ^*					81.88	76.80	65.93	43.21						
		σ^*					23.27	16.55	10.44	4.92						
	5	μ^*						74.54	65.86							
		σ^*						10.41	4.90							

14	0	μ^*	100.0	98.53	96.11	92.53	87.54	80.87	72.31	61.85	49.78	36.85	24.21	13.18	4.87
		σ^*	100.0	89.00	78.95	69.59	60.79	52.48	44.59	37.10	29.98	23.22	16.82	10.78	5.14
	1	μ^*		97.39	95.34	92.14	87.45	80.86	72.03	60.79	47.45	33.01	19.12	7.65	
		σ^*		78.28	68.56	59.55	51.12	43.20	35.74	28.71	22.09	15.88	10.09	4.77	
	2	μ^*			93.79	91.16	87.03	80.82	71.92	59.86	44.82	28.08	12.19		
		σ^*			59.19	50.57	42.55	35.06	28.06	21.52	15.43	9.79	4.62		
	3	μ^*				89.29	86.00	80.55	71.90	58.92	41.12	20.03			
		σ^*				42.35	34.76	27.73	21.21	15.18	9.62	4.55			
	4	μ^*					83.86	79.70	71.88	57.75	33.95				
		σ^*					27.63	21.08	15.05	9.53	4.51				
	5	μ^*						77.44	71.64	55.56					
		σ^*						15.01	9.49	4.48					
	6	μ^*							69.91						
		σ^*							4.48						

Table III (continued)

n	r_1 \ r_2		0	1	2	3	4	5	6	7	8	9	10	11	12	13
15	0	μ^*	100.0	98.68	96.52	93.40	89.11	83.42	76.15	67.21	56.70	45.04	32.97	21.50	11.67	4.33
		σ^*	100.0	89.72	80.30	71.51	63.21	55.35	47.88	40.76	33.97	27.49	21.33	15.47	9.93	4.74
	1	μ^*		97.62	95.79	92.99	88.97	83.42	76.04	66.62	55.23	42.36	29.03	16.65	6.64	
		σ^*		79.69	70.54	62.04	54.06	46.54	39.42	32.68	26.30	20.28	14.60	9.30	4.40	
	2	μ^*			94.37	92.04	88.50	83.31	76.02	66.22	53.79	39.28	24.07	10.30		
		σ^*			61.70	53.53	45.89	38.73	32.00	25.68	19.73	14.18	9.01	4.27		
	3	μ^*				90.33	87.46	82.93	76.01	65.93	52.12	34.94	16.38			
		σ^*				45.70	38.43	31.66	25.33	19.43	13.93	8.85	4.20			
	4	μ^*					85.49	81.95	75.86	65.73	49.75	26.97				
		σ^*					31.55	25.18	19.27	13.80	8.76	4.15				
	5	μ^*						79.80	75.25	65.61	44.82					
		σ^*						19.22	13.74	8.71	4.13					
	6	μ^*							73.19	65.56						
		σ^*							8.70	4.12						

normal distribution. In this figure, the graphs for $r_1 = 0, 1, \cdots, 12$ show a parallelism as r_1 changes. Thus, for any corresponding value of r_2 the efficiency decreases by about the same amount for each change in the value of r_1.

(3) Using Table III again and reading the entries for σ^* in diagonal fashion, one can see that, for a given n and fixed uncensored sample size ($r_1 + r_2 =$ constant), the efficiency of the best estimate of σ *is remarkably constant independently of how r_1 and r_2 are chosen.* In other words, there is practically no difference in efficiency irrespective of the proportion of the relative censoring *from either side.*

This can be observed very clearly in Fig. 2. The approximate horizontal lines show constancy of the relative efficiencies of σ^* for the known elements (k) of the sample whatever may be the individual values of r_1 and r_2.

(4) From Table III (and graphs similar to Fig. 2), one can construct the following table showing how the efficiency in estimating σ^* varies with the number of uncensored values for each sample size to serve as a rough guide in censoring.

Rough guide for assessing approximate efficiency (per cent) * *of estimate of* σ

Sample Size n	Number of uncensored observations in sample, or $k = n - (r_1 + r_2)$													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
11	6	13	21	30	39	50	61	73	86	100				
12	6	12	19	27	35	44	54	64	75	87	100			
13	5	11	17	24	31	39	48	57	67	77	88	100		
14	5	10	15	22	28	36	43	51	59	69	78	89	100	
15	4	9	14	20	26	32	39	46	54	62	71	80	90	100

* These values are rounded averages of different combinations of censoring and are within 2 or 3 per cent in almost all cases

The information in this rough guide for censoring is also illustrated in Fig. 3 The efficiency in estimating σ^* for varying proportions of censoring is shown in the graph for samples of size 10, 15, 20 and large n. The latter value was obtained from Gupta [1] and represented single censoring. However, as stated previously, the efficiency in estimating σ^* depends primarily upon the proportion of uncensored elements irrespective of the side and can be used in this way.

(5) Figs. 4 and 5 show the efficiencies of the alternative estimate from a sample of size 15 relative to the correspondingly best linear estimate for the mean and the standard deviation respectively.

Judging from these figures, the worst efficiencies of the alternative estimate for estimating both the mean and standard deviations are attained for singly censored samples (*i.e., only* r_1 *or* $r_2 = 0$). Thus, the alternative estimate is relatively more precise when applied to doubly censored samples. The alternative estimate was proposed by judging the results of a comparison of efficiencies using a singly

Table IV

...iances and Relative Efficiencies of Alternative Estimates of the Mean ($\mu^{*\prime}$) and Standard Deviation ($\sigma^{*\prime}$) for Censored Samples of Size 12 and 15 from a Normal Population

	0	1	2	3	4	5	6	7	8	9	10	11	12	1
*') V.	.08333333	.08532138	.08921286	.09569251	.10617803	.12326618	.15197055	.20294226	.30233969	.53167098	1.30441131			
R.E.	100.00	99.51	98.38	96.72	94.59	92.13	89.71	87.90	87.64	90.44	100.00			
*') V.	.04693011	.05541733	.06584693	.07855768	.09459627	.11566732	.14474943	.18762545	.25730631	.39089620	.76011855			
R.E.	99.89	97.03	94.20	92.04	90.42	89.29	88.66	88.66	89.60	92.36	100.00			
*') V.		.08688517	.08998634	.09535109	.10440203	.11990788	.14777036	.20261677	.33120159	.78642201				
R.E.		99.11	98.38	97.39	96.20	95.03	94.25	94.50	96.48	100.00				
') V.		.06478228	.07715733	.09281482	.11337921	.14183822	.18414527	.25419403	.39403316	.81947829				
R.E.		96.77	95.48	94.58	94.15	94.19	94.75	95.93	97.81	100.00				
') V.			.09205743	.09613108	.10369458	.11794105	.14676609	.21537209	.46137626					
R.E.			98.13	97.67	97.11	96.68	96.85	98.21	100.00					
') V.			.09299998	.11391167	.14281363	.18574083	.25691562	.39999265	.84501770					
R.E.			95.09	94.97	95.30	96.14	97.51	99.28	100.00					
') V.				.09854217	.10405259	.11637233	.14734277	.26167710						
R.E.				97.77	97.79	97.99	98.85	100.00						
') V.				.14326687	.18682141	.25894808	.40403858	.85814724						
R.E.				95.53	96.54	98.04	99.80	100.00						
') V.					.10664702	.11546756	.15239128							
R.E.					98.38	99.08	100.00							
') V.					.25964368	.40599317	.86468798							
R.E.					98.17	99.98	100.00							
') V.						.11751619								
R.E.						100.00								
*') V.						.86669069								
R.E.						100.00								

*') V.	.06676667	.06703620	.06826743
R.E.	100.00		

0	$(\mu*l)$ V.	.06676667	.06783620	.06996743	.07323737	.07806417	.08512564	.09551426	.11105404	.13495703	.17330497	.23882285	.36232523	.63848625	1.54037885
	R.E.	100.00	99.60	98.71	97.46	95.84	93.88	91.66	89.32	87.12	85.40	84.66	85.58	89.47	100.00
	$(\sigma*l)$ V.	.03667132	.04199425	.04825549	.05545771	.06393382	.07415317	.08679542	.10290237	.12417377	.15360043	.19699885	.26744195	.40210996	.77230522
	R.E.	99.90	97.23	94.53	92.38	90.64	89.25	88.15	87.35	86.85	86.75	87.19	88.54	91.74	100.00
1	$(\mu*l)$ V.		.06883025	.07060707	.07337808	.07754855	.08378381	.09319554	.10773719	.13110231	.17103630	.24649116	.41734709	1.00373313	
	R.E.		99.22	98.57	97.70	96.63	95.38	94.08	92.88	92.06	92.01	93.17	95.95	100.00	
	$(\sigma*l)$ V.		.04757932	.05459531	.06286426	.07280110	.08506272	.10067027	.12130831	.14998813	.19270808	.26346042	.40451706	.83217195	
	R.E.		96.62	95.12	93.93	93.08	92.54	92.31	92.40	92.86	93.75	95.21	97.38	100.00	
2	$(\mu*l)$ V.			.07192272	.07413157	.07764224	.08314965	.09187723	.10612952	.13071264	.17732697	.28290628	.64711601		
	R.E.			98.23	97.70	97.02	96.23	95.45	94.86	94.81	95.72	97.89	100.00		
	$(\sigma*l)$ V.			.06297264	.07309736	.08556985	.10141243	.12232669	.15136196	.19461599	.26642265	.41072321	.85860349		
	R.E.			94.28	93.63	93.28	93.26	93.57	94.26	95.38	96.99	98.98	100.00		
3	$(\mu*l)$ V.				.07566586	.07837611	.08298324	.09085279	.10482738	.13179671	.19342211	.40692760			
	R.E.				97.54	97.25	96.88	96.54	96.46	97.05	98.65	100.00			
	$(\sigma*l)$ V.				.08580912	.10194081	.12319669	.15265533	.19648232	.26920271	.41556039	.87312546			
	R.E.				93.42	93.50	93.93	94.73	95.97	97.66	99.58	100.00			
4	$(\mu*l)$ V.					.08009668	.08351673	.09014086	.10361683	.13537627	.24716376				
	R.E.					97.36	97.40	97.49	97.89	98.99	100.00				
	$(\sigma*l)$ V.					.12350319	.15333514	.19764253	.27106519	.41878810	.88168661				
	R.E.					94.02	94.90	96.20	97.94	99.85	100.00				
5	$(\mu*l)$ V.						.08536772	.09008179	.10250554	.14874275					
	R.E.						97.86	98.35	99.13	100.00					
	$(\sigma*l)$ V.						.19802892	.27197904	.42061691	.88658942					
	R.E.						96.26	98.05	99.97	100.00					
	$(\mu*l)$ V.							.09184730	.10169465						
	R.E.							99.17	100.00						
	$(\sigma*l)$ V.							[illegible]	[illegible]						

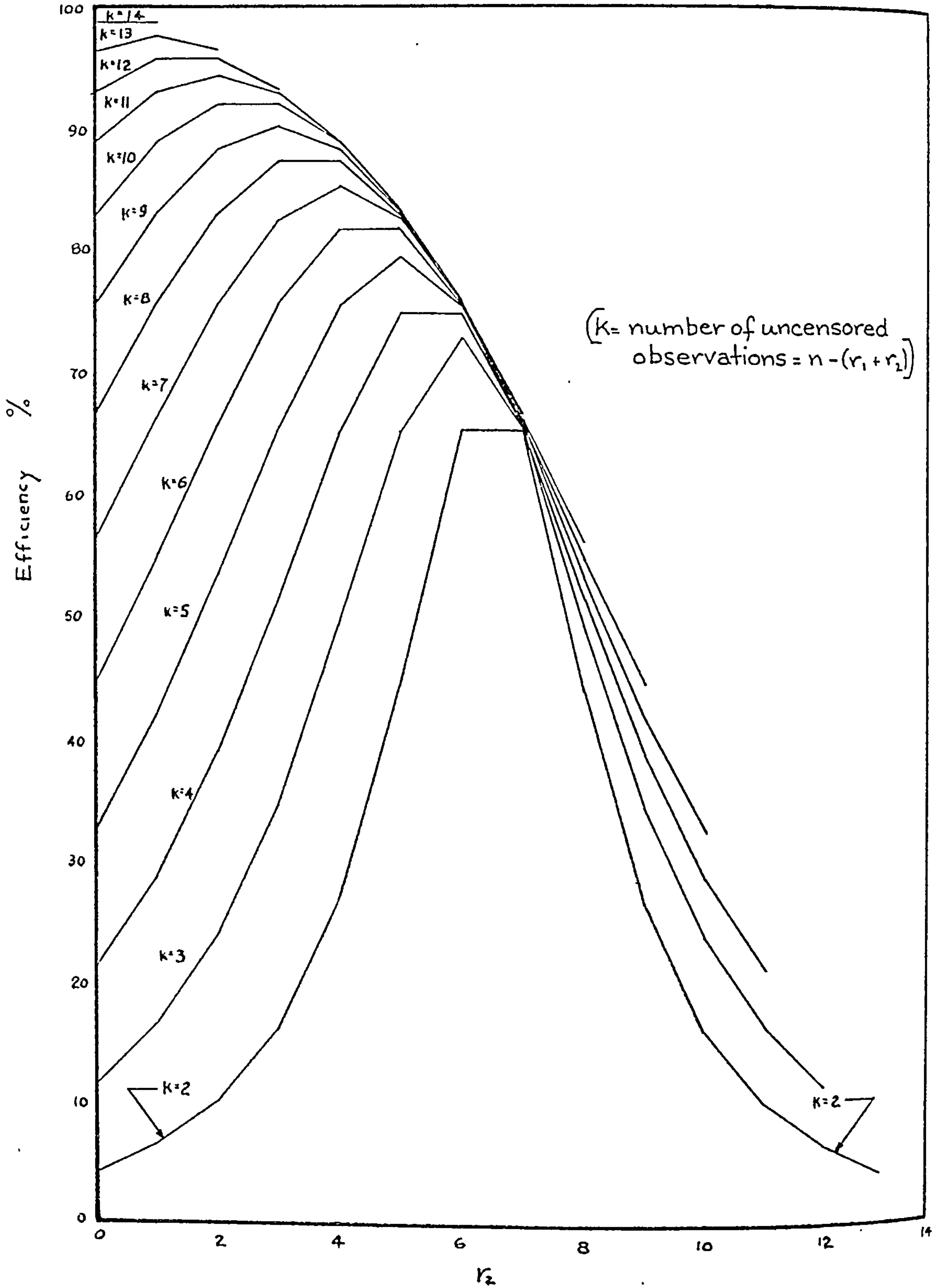


FIG. 1. Relative efficiency of the best linear estimate of μ under all conditions of censoring a sample of size 15 from the normal distribution.

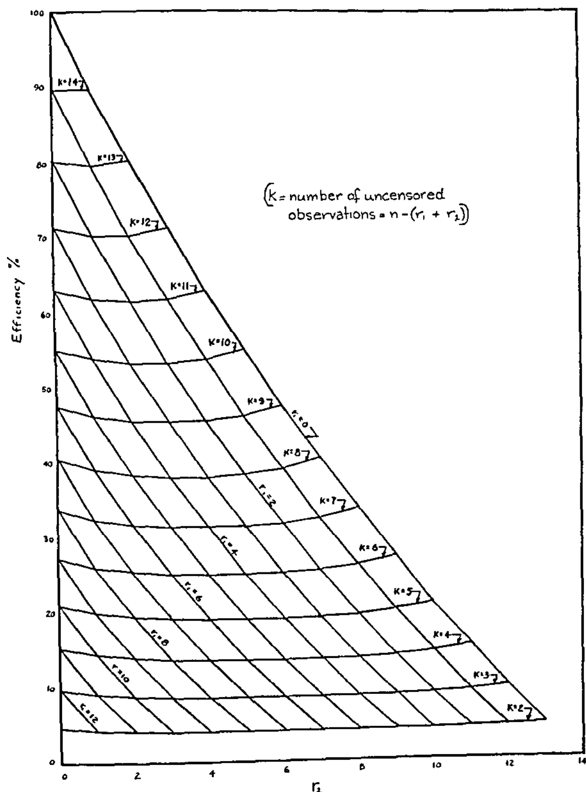


FIG. 2. Relative efficiency of the best linear estimate of σ under all conditions of censoring a sample of size 15 from the normal distribution.

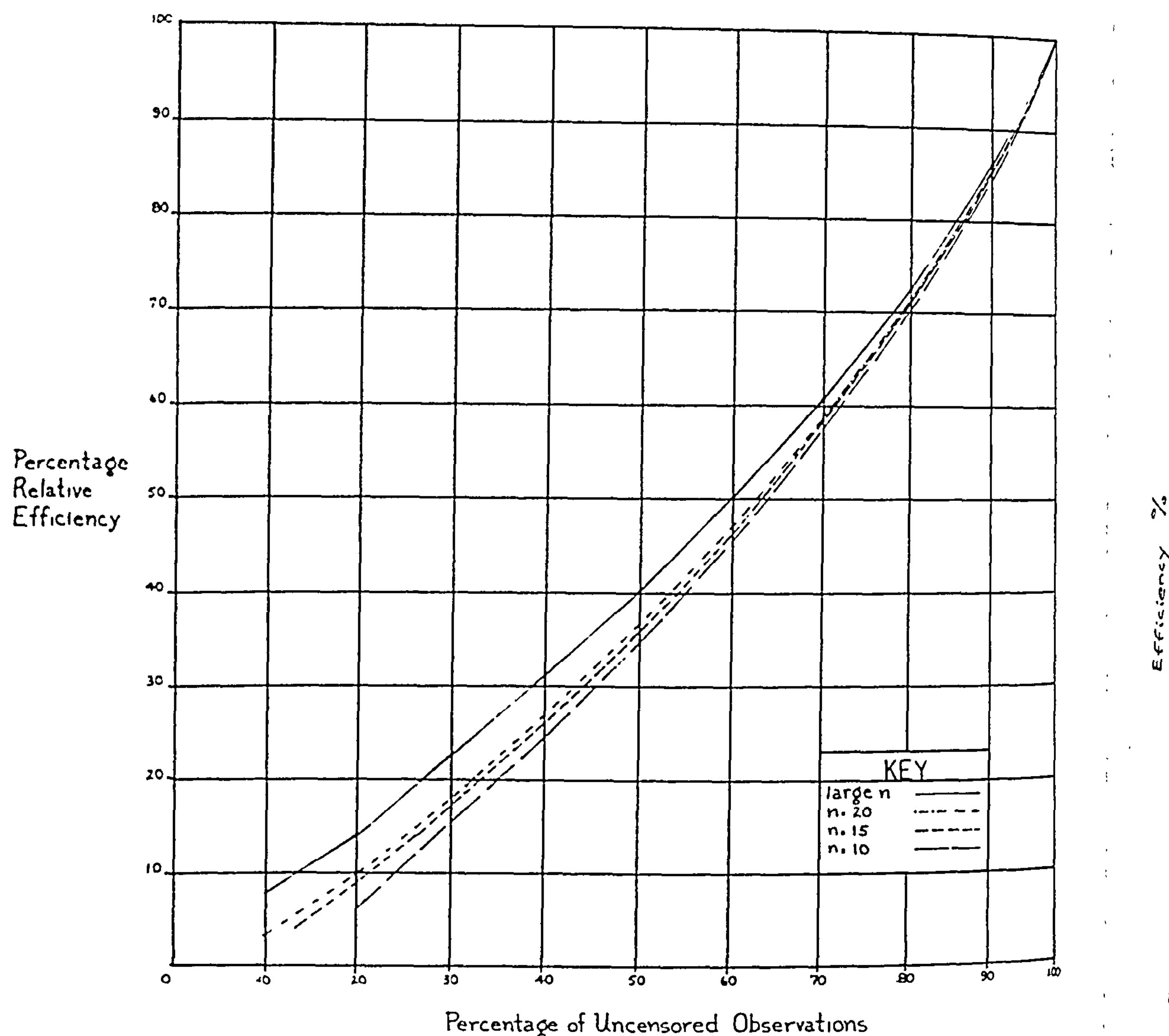


FIG. 3. Approximate efficiencies in estimating σ^* for censored samples of size 10, 15, 20 and large n.

censored sample. The present graphs show that the alternative estimate is even better than previously supposed.

Also, one can observe that for $r_1 = 0$, the alternative estimate of σ is more efficient than the corresponding estimate of the mean. In fact for other values of r_1, the efficiencies are much more concentrated for the former than for the latter. Again, the drop in efficiency for the estimate of the mean is much sharper than that for the standard deviation. This shows that the alternative estimate also appears better if one judges its value by considering its efficiency in estimation of the standard deviation rather than of the mean alone.

Addendum. The extension of Tables I, II, and III of this paper are now available for $16 \leqq n \leqq 20$ in eight decimal places. Copies may be obtained from the authors.

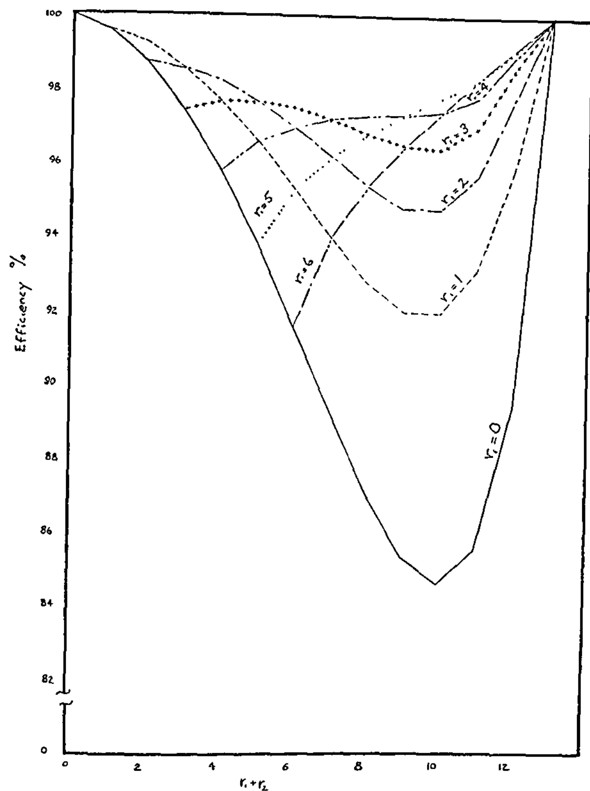


Fig. 4 Relative efficiency of the alternative estimate of μ under conditions of censoring a sample of size 15 from the normal population.

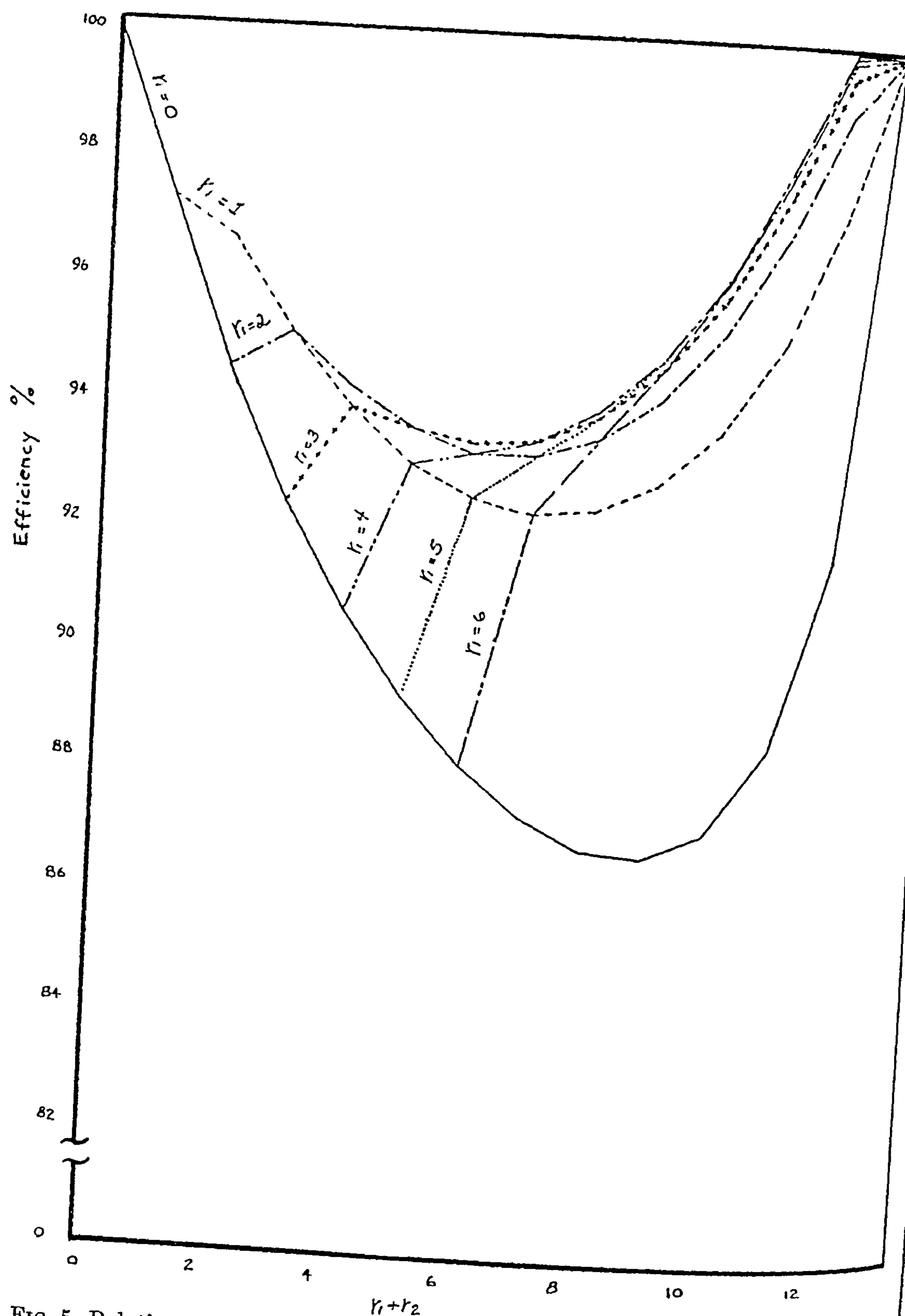


FIG. 5. Relative efficiency of the alternative estimate of σ under conditions of censoring a sample of size 15 from the normal population.

REFERENCES

[1] A. K. Gupta, "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, Vol. 39 (1952), pp. 88–95.

[2] A. E. Sarhan and B. G. Greenberg, "Estimation of location and scale parameters by order statistics from singly and doubly censored samples. Part I The normal distribution up to samples of size 10," *Ann. Math. Stat.*, Vol. 27, No. 2 (1956) pp. 427–451.

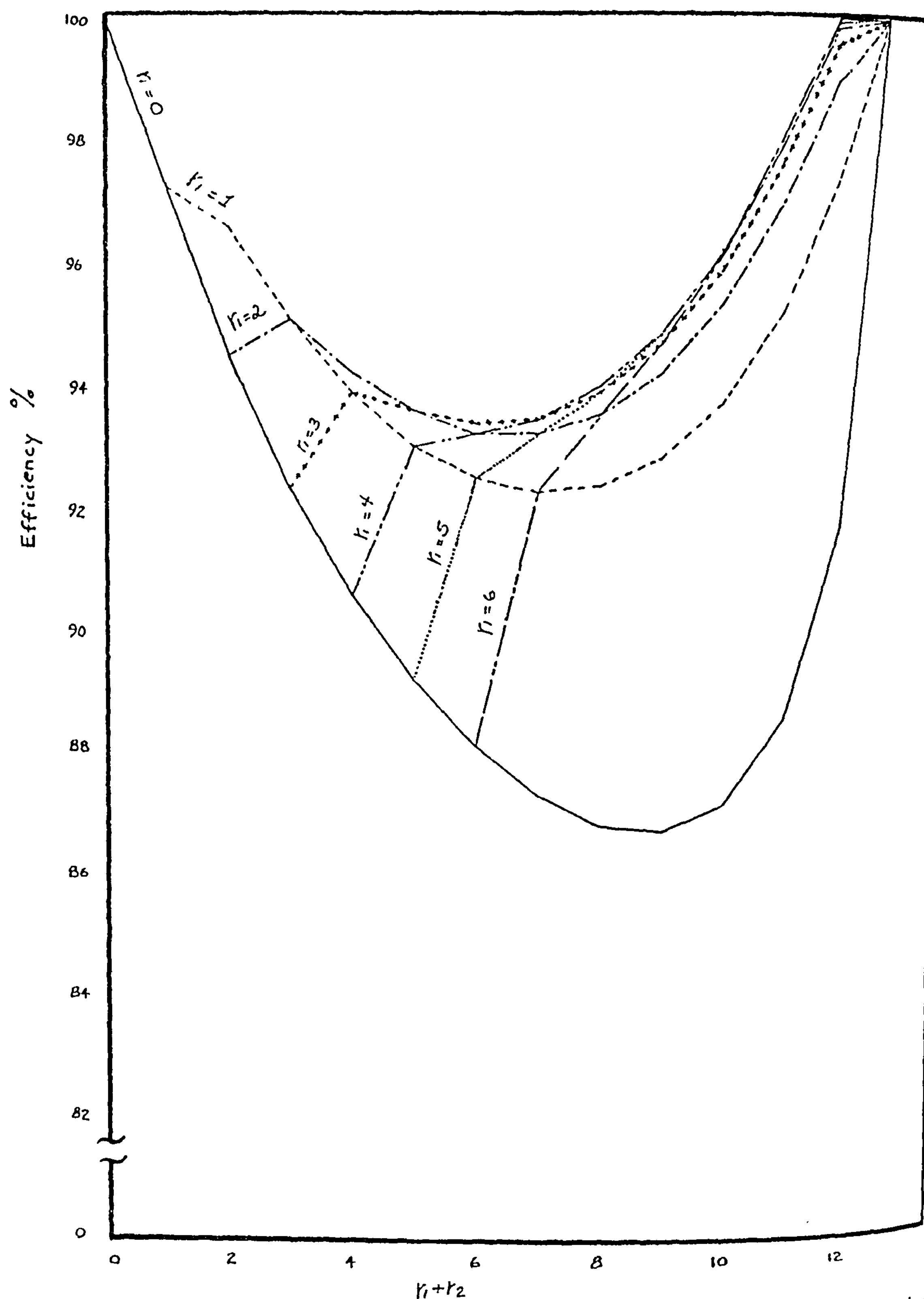


FIG. 5. Relative efficiency of the alternative estimate of σ under conditions of censoring a sample of size 15 from the normal population.

REFERENCES

[1] A. K. GUPTA, "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, Vol. 39 (1952), pp. 88–95.

[2] A. E. SARHAN AND B. G. GREENBERG, "Estimation of location and scale parameters by order statistics from singly and doubly censored samples. Part I. The normal distribution up to samples of size 10," *Ann. Math Stat*, Vol. 27, No. 2 (1956) pp. 427–451.

GENERALIZATIONS OF A GAUSSIAN THEOREM

By Paul S. Dwyer

University of Michigan

1. Introduction and Summary. Plackett [1] has discussed the history and generalizations of the Gaussian theorem which states that least squares estimates are linear unbiased estimates with minimum variance. General forms of the theorem are due to Aitken [2], [3] and Rao [4], [5]. The essence of the proof for Aitken's general case consists in minimizing, simultaneously, certain quadratic forms involving linear combinations of the parameters. Plackett derived Aitken's result by using a matrix relation. The proof of the theorem follows quickly once the relation is established. A somewhat similar but simpler matrix relation is used by Rao ([4], page 10).

Aitken [2] and Rao [4], [5] obtain minimum variance with the use of Lagrange multipliers. Unless one has a method of working with matrices of derivatives it seems necessary to differentiate with respect to the many scalars constituting the matrices and to assemble the results in desired matrix form. Authors frequently give only the assembled results ([4], page 10, [5], page 17, [6], page 83).

The question arises as to whether it is possible to use the logically preferable matrix derivative methods of minimization. It is shown below that the use of matrices of partial derivatives [7] leads logically to the solution without the necessity of changing to and from scalar notation, or without the necessity of establishing some relation which implicitly contains the solution. Matrix derivative methods seem to be preferable methods for undertaking solutions of problems of simultaneous matrix minimization with side conditions for the same reason that derivative methods are preferable to the use of some (unknown) relation in solving problems of minimization involving scalars. They may also be used in establishing the relation which may then be verified without their use.

The paper includes generalizations of the results of Aitken [2], [3], Rao [4], [5], and David and Neyman [8]. It gives a general formula for simultaneous unbiased estimators of linear functions of parameters when the parameters are subject to linear restrictions and shows how the results are applicable to special cases. It provides formulas for the variance matrix of these estimators. It generalizes a matrix relation used by Plackett [1]. It uses the matrix square root transformation in establishing the general result for the variance of (weighted) residuals when there may be linear restrictions on the parameters. It provides a generalization of a formula of David and Neyman [8] in estimating the variance matrix of the unbiased linear estimators.

2. The least squares solution. The (inconsistent) observational equations are

$$A\theta = x \tag{2.1}$$

Received February 18, 1957; revised September 17, 1957.

and the true linear regression is given by

$$\mathcal{E}(x) = A\theta, \tag{2.2}$$

where the values of x, A, and θ are real. We set

$$A\theta - x = \epsilon \tag{2.3}$$

so that

$$\mathcal{E}(\epsilon) = 0. \tag{2.4}$$

In determining the least squares regression we have $\theta(s \times 1)$ as the vector of unknown parameters, $x(n \times 1)$ as the vector of measurements of the variable of regression, $\epsilon(n \times 1)$ as the vector of errors and $A(n \times s)$ as the matrix of measurements of the regressed variables We take $s < n$ and A of rank s. Further, under the usual regression condition of fixed A,

$$\begin{aligned} V &= \mathcal{E}(xx^T) - \mathcal{E}(x)\mathcal{E}(x^T) = \mathcal{E}(\epsilon\epsilon^T) = V^T \\ &= \text{var }(x) = \text{var }\epsilon \end{aligned} \tag{2.5}$$

is the dispersion matrix of x and ϵ. We limit our discussion to the case where V is positive definite. A common dimensionless generalization of the least squares concept uses weights for the observations with $W = V^{-1}$ and leads to

$$Q = \epsilon^T V^{-1} \epsilon = (A\theta - x)^T V^{-1}(A\theta - x) \tag{2.6}$$

as the form to be minimized. The value of θ which minimizes (2.6) is known to be

$$\theta^* = (A^T V^{-1} A)^{-1} A^T V^{-1} x. \tag{2.7}$$

This result can be derived using symbolic matrix derivatives ([7], page 524). We have successively

$$Q = \theta^T A^T V^{-1} A\theta - \theta^T A^T V^{-1} x - x^T V^{-1} A\theta + x^T V^{-1} x, \tag{2.8}$$

$$\frac{\partial Q}{\partial\langle\theta\rangle} = J^T A^T V^{-1} A\theta + \theta^T A^T V^{-1} AJ - J^T A^T V^{-1} x - x^T V^{-1} AJ, \tag{2.9}$$

$$\frac{\partial\langle Q\rangle}{\partial\theta} = A^T V^{-1} A\theta K^T + A^T V^{-1} A\theta K - A^T V^{-1} x K^T - A^T V^{-1} x K, \tag{2.10}$$

and since Q is scalar, $K = K^T = 1$. Setting $\theta = \theta^*$ when $\dfrac{\partial\langle Q\rangle}{\partial\theta} = 0$, we get (2.7).

We note that θ^* is an unbiased estimate of θ because of (2.2) and (2.7)

3. Linear estimates with minimum variance. Now consider the k linear parametric functions

$$\phi = L\theta, \tag{3 1}$$

GENERALIZATIONS OF A GAUSSIAN THEOREM

By Paul S. Dwyer

University of Michigan

1. Introduction and Summary. Plackett [1] has discussed the history generalizations of the Gaussian theorem which states that least squares estim are linear unbiased estimates with minimum variance. General forms of theorem are due to Aitken [2], [3] and Rao [4], [5]. The essence of the proo Aitken's general case consists in minimizing, simultaneously, certain quad forms involving linear combinations of the parameters. Plackett derived Aitl result by using a matrix relation. The proof of the theorem follows quickly the relation is established. A somewhat similar but simpler matrix relation is by Rao ([4], page 10).

Aitken [2] and Rao [4], [5] obtain minimum variance with the use of Lagr multipliers. Unless one has a method of working with matrices of derivativ seems necessary to differentiate with respect to the many scalars constituting matrices and to assemble the results in desired matrix form. Authors freque give only the assembled results ([4], page 10, [5], page 17, [6], page 83).

The question arises as to whether it is possible to use the logically prefe matrix derivative methods of minimization. It is shown below that the u matrices of partial derivatives [7] leads logically to the solution without necessity of changing to and from scalar notation, or without the necessit establishing some relation which implicitly contains the solution. Matrix de tive methods seem to be preferable methods for undertaking solutions of p lems of simultaneous matrix minimization with side conditions for the s reason that derivative methods are preferable to the use of some (unkn relation in solving problems of minimization involving scalars. They may be used in establishing the relation which may then be verified without their

The paper includes generalizations of the results of Aitken [2], [3], Rao [5], and David and Neyman [8]. It gives a general formula for simultan unbiased estimators of linear functions of parameters when the parameter subject to linear restrictions and shows how the results are applicable to sp cases. It provides formulas for the variance matrix of these estimators. It eralizes a matrix relation used by Plackett [1]. It uses the matrix square transformation in establishing the general result for the variance of (weigh residuals when there may be linear restrictions on the parameters. It provic generalization of a formula of David and Neyman [8] in estimating the vari matrix of the unbiased linear estimators.

2. The least squares solution. The (inconsistent) observational equation

$$A\theta = x \tag{2.1}$$

Received February 18, 1957; revised September 17, 1957.

and the true linear regression is given by

$$\mathcal{E}(x) = A\theta, \tag{2.2}$$

where the values of x, A, and θ are real. We set

$$A\theta - x = \epsilon \tag{2.3}$$

so that

$$\mathcal{E}(\epsilon) = 0. \tag{2.4}$$

In determining the least squares regression we have $\theta(s \times 1)$ as the vector of unknown parameters, $x(n \times 1)$ as the vector of measurements of the variable of regression, $\epsilon(n \times 1)$ as the vector of errors and $A(n \times s)$ as the matrix of measurements of the regressed variables. We take $s < n$ and A of rank s. Further, under the usual regression condition of fixed A,

$$\begin{aligned} V &= \mathcal{E}(xx^T) - \mathcal{E}(x)\mathcal{E}(x^T) = \mathcal{E}(\epsilon\epsilon^T) = V^T \\ &= \text{var}\,(x) = \text{var}\,\epsilon \end{aligned} \tag{2.5}$$

is the dispersion matrix of x and ϵ. We limit our discussion to the case where V is positive definite. A common dimensionless generalization of the least squares concept uses weights for the observations with $W = V^{-1}$ and leads to

$$Q = \epsilon^T V^{-1}\epsilon = (A\theta - x)^T V^{-1}(A\theta - x) \tag{2.6}$$

as the form to be minimized. The value of θ which minimizes (2.6) is known to [illegible]

$$\theta^* = (A^T V^{-1} A)^{-1} A^T V^{-1} x. \tag{2.7}$$

This result can be derived using symbolic matrix derivatives [illegible]
We have successively

$$Q = \theta^T A^T V^{-1} A\theta - \theta^T A^T V^{-1} x - x^T V^{-1} A\theta + [illegible] \tag{2.8}$$

$$\frac{\partial Q}{\partial\langle\theta\rangle} = J^T A^T V^{-1} A\theta + \theta^T A^T V^{-1} A J - J^T A^T [illegible] \tag{2.9}$$

$$\frac{\partial\langle Q\rangle}{\partial\theta} = A^T V^{-1} A\theta K^T + A^T V^{-1} A\theta K - A^T [illegible] \tag{2.10}$$

and since Q is scalar, $K = K^T = 1$. Setting $\theta =$ [illegible]

We note that θ^* is an unbiased estimate of [illegible]

3. Linear estimates with minimum [illegible] ametric functions

$$\phi = [illegible] \tag{3.1}$$

where $L = L(k \times s)$ is known. Then $\phi = \phi(k \times 1)$. We wish to find

$$\phi^* = L\theta^* \tag{3.2}$$

such that ϕ^* is an unbiased estimate of ϕ with minimum variance. This means that the diagonal terms of var ϕ (a matrix of order $k \times k$) should attain their minimum values simultaneously. Following Aitken we consider solutions of the form

$$\phi^* = Bx \tag{3.3}$$

and determine $B = B(k \times n)$. Rao [4] has shown that this homogeneous form is the general form. The relation

$$(BA - L)\theta = 0 \tag{3.4}$$

follows from (3.3), (2.2), and (3.1) in accordance with the requirement that ϕ^* be an unbiased estimate of ϕ.

Aitken [3] has shown using Lagrangian multipliers and Plackett [1] using a matrix relation that the value of θ^* in (3.2) which minimizes the diagonal term of var Q^* is identical with the θ^* resulting from least squares as given by (2.7). This Aitken theorem is a generalization of the Gaussian theorem that least squares linear estimators are unbiased with minimum variance.

Rao [5] further generalized the theorem with a consideration of linear restrictions on the parameters when $k = 1$. The argument is given below for the more general k. The preparation of the problem for minimization is similar to that of Rao in the special case with $k = 1$, though there are some modifications. The $u < s$ independent linear restrictions may be indicated by

$$g = R\theta \equiv 0, \tag{3.5}$$

where $R = R(u \times s)$ and $g = g(u \times 1)$, without loss of generality since any term not having some θ_i as a factor may be multiplied by $\theta_0 = 1$ and s replaced by $s' = s + 1$. We premultiply by the undetermined $D = D(k \times u)$ to get

$$DR\theta = Dg, \tag{3.6}$$

in which the matrix coefficient of θ has the same order as BA and L. Then the condition for unbiased estimation *and* the specific side conditions are incorporated in the matrix relation

$$(L - BA)\,\theta \equiv 0 \equiv DR\theta - Dg \tag{3.7}$$

so that the conditions for estimation can be written in the form

$$(L - BA - DR) = 0 \quad \text{and} \quad Dg = 0. \tag{3.8}$$

Specifically our purpose is the minimization of the diagonal terms of var ϕ^* subject to (3.8). Now

$$\text{var}\,\phi^* = \mathcal{E}(\phi^*\phi^{*T}) - \mathcal{E}(\phi^*)\mathcal{E}(\phi^{*T}) = B[\mathcal{E}(xx^T) - \mathcal{E}(x)\mathcal{E}(x^T)]B^T = BVB.^T \tag{3.9}$$

e can then use

10) $$\psi = BVB^T + 2(L - BA - DR)\Lambda + 2Dg\mu^T,$$

ie $\psi = \psi\ (k \times k)$, $\Lambda = \Lambda\ (s \times k)$, and $\mu = \mu\ (k \times 1)$ and differentiate with ...t to B and D. We have

$$\frac{\partial \psi}{\partial \langle B \rangle} = JVB^T + BVJ^T - 2JA\Lambda,$$

$$\frac{\partial \langle \psi \rangle}{\partial B} = KBV + K^TBV - 2K\Lambda^TA^T,$$

o that the critical value, for each and every diagonal term, is given by

3.11) $$BV = \Lambda^TA^T.$$

Again

$$\frac{\partial \psi}{\partial \langle D \rangle} = -2JR\Lambda + 2Jg\mu^T,$$

$$\frac{\partial \langle \psi \rangle}{\partial D} = -2K\Lambda^TR^T + 2K\mu g^T,$$

$$\frac{\partial \psi_{ii}}{\partial D} = -2K_{ii}\Lambda^TR^T + 2K_{ii}\mu g^T,$$

that, for each and every diagonal term

$$\Lambda^TR^T = \mu g^T,$$

o by (3.5),

(3.12) $$\Lambda^TR^T = 0.$$

From (3.11) we get

(3.13) $$B = \Lambda^TA^TV^{-1}.$$

... in the first equation of (3.8), we arrive at

(3.14) $$\Lambda^TA^TV^{-1}A + DR = L.$$

This equation and (3.12), for the special case with $k = 1$, were derived and emphasized by Rao [5], [17].

We next derive an estimate of ϕ in terms of Λ^T and θ^* for general k. We just multiply (3.14) by θ^* and use $R\theta^* = 0$ to get

(3.15) $$\Lambda^TA^TV^{-1}A\theta^* = \phi^*.$$

The corresponding estimate in terms of Λ^T and x is

(3.16) $$\phi^* = Bx = \Lambda^TA^TV^{-1}x.$$

where $L = L(k \times s)$ is known. Then $\phi = \phi(k \times 1)$. We wish to find

$$\phi^* = L\theta^* \tag{3.2}$$

such that ϕ^* is an unbiased estimate of ϕ with minimum variance. This means that the diagonal terms of var ϕ (a matrix of order $k \times k$) should attain their minimum values simultaneously. Following Aitken we consider solutions of the form

$$\phi^* = Bx \tag{3.3}$$

and determine $B = B(k \times n)$. Rao [4] has shown that this homogeneous form is the general form. The relation

$$(BA - L)\theta = 0 \tag{3.4}$$

follows from (3.3), (2.2), and (3.1) in accordance with the requirement that ϕ^* be an unbiased estimate of ϕ.

Aitken [3] has shown using Lagrangian multipliers and Plackett [1] using a matrix relation that the value of θ^* in (3.2) which minimizes the diagonal term of var Q^* is identical with the θ^* resulting from least squares as given by (2.7). This Aitken theorem is a generalization of the Gaussian theorem that least squares linear estimators are unbiased with minimum variance.

Rao [5] further generalized the theorem with a consideration of linear restrictions on the parameters when $k = 1$. The argument is given below for the more general k. The preparation of the problem for minimization is similar to that of Rao in the special case with $k = 1$, though there are some modifications. The $u < s$ independent linear restrictions may be indicated by

$$g = R\theta \equiv 0, \tag{3.5}$$

where $R = R(u \times s)$ and $g = g(u \times 1)$, without loss of generality since any term not having some θ_i as a factor may be multiplied by $\theta_0 = 1$ and s replaced by $s' = s + 1$. We premultiply by the undetermined $D = D(k \times u)$ to get

$$DR\theta = Dg, \tag{3.6}$$

in which the matrix coefficient of θ has the same order as BA and L. Then the condition for unbiased estimation *and* the specific side conditions are incorporated in the matrix relation

$$(L - BA)\,\theta \equiv 0 \equiv DR\theta - Dg \tag{3.7}$$

so that the conditions for estimation can be written in the form

$$(L - BA - DR) = 0 \quad \text{and} \quad Dg = 0. \tag{3.8}$$

Specifically our purpose is the minimization of the diagonal terms of var ϕ^* subject to (3.8). Now

$$\text{var}\,\phi^* = \mathcal{E}(\phi^*\phi^{*T}) - \mathcal{E}(\phi^*)\mathcal{E}(\phi^{*T}) = B[\mathcal{E}(xx^T) - \mathcal{E}(x)\mathcal{E}(x^T)]B^T = BVB^T. \tag{3.9}$$

We can then use

$$\psi = BVB^T + 2(L - BA - DR)\,\Lambda + 2Dg\mu^T, \tag{3.10}$$

where $\psi = \psi\ (k \times k)$, $\Lambda = \Lambda\ (s \times k)$, and $\mu = \mu\ (k \times 1)$ and differentiate with respect to B and D. We have

$$\frac{\partial\psi}{\partial\langle B\rangle} = JVB^T + BVJ^T - 2JA\Lambda,$$

$$\frac{\partial\langle\psi\rangle}{\partial B} = KBV + K^TBV - 2K\Lambda^TA^T,$$

so that the critical value, for each and every diagonal term, is given by

$$BV = \Lambda^TA^T. \tag{3.11}$$

Again

$$\frac{\partial\psi}{\partial\langle D\rangle} = -2JR\Lambda + 2Jg\mu^T,$$

$$\frac{\partial\langle\psi\rangle}{\partial D} = -2K\Lambda^TR^T + 2K\mu g^T,$$

$$\frac{\partial\psi_{ii}}{\partial D} = -2K_{ii}\Lambda^TR^T + 2K_{ii}\mu g^T,$$

so that, for each and every diagonal term

$$\Lambda^TR^T = \mu g^T,$$

so by (3.5),

$$\Lambda^TR^T = 0. \tag{3.12}$$

From (3.11) we get

$$B = \Lambda^TA^TV^{-1}. \tag{3.13}$$

Substituting in the first equation of (3.8), we arrive at

$$\Lambda^TA^TV^{-1}A + DR = L. \tag{3.14}$$

This equation and (3.12), for the special case with $k = 1$, were derived an emphasized by Rao [5], [17].

We next derive an estimate of ϕ in terms of Λ^T and θ^* for general k. We ju multiply (3.14) by θ^* and use $R\theta^* = 0$ to get

$$\Lambda^TA^TV^{-1}A\theta^* = \phi^*. \tag{3.15}$$

The corresponding estimate in terms of Λ^T and x is

$$\phi^* = Bx = \Lambda^TA^TV^{-1}x. \tag{3.16}$$

It follows that θ^* satisfies

$$\Lambda^T A^T V^{-1} A \theta^* = \Lambda^T A^T V^{-1} x. \tag{3.17}$$

Equations (3.17) and (3.12) may be considered to be basic relations in θ^* an Λ^T.

4. The general Gaussian theorem. We next demonstrate the general Gaussia theorem that the value of θ^* obtained by least squares is consistent with that (3.17) and (3.12). We note first that θ^* in the general solution is an $s \times 1$ vect and that the general solution is obtained by premultiplying θ^* by the fixe $k \times s$ matrix L. The general theorem is established by proving the typical ca with $k = 1$ so that L, B, D, and Λ are vectors and (3.17) becomes

$$\lambda^T A^T V^{-1} A \theta^* = \lambda^T A^T V^{-1} x, \tag{4.1}$$

where λ^T is $\lambda^T(1 \times s)$. Also (3.12) becomes

$$\lambda^T R^T = 0. \tag{4.2}$$

Now we wish to minimize the scalar $Q = \epsilon^T V^{-1} \epsilon$, subject to the restriction co ditions. Then

$$Q' = (A\theta - x)^T V^{-1}(A\theta - x) + 2(l - bA - dR)\lambda + 2\gamma R\theta. \tag{4.3}$$

Differentiation with respect to θ and d leads to the "normal" equations

$$A^T V^{-1} A \theta^* - A^T V^{-1} x + R^T \gamma^T = 0, \tag{4.4}$$

$$\lambda^T R^T = 0. \tag{4.5}$$

Premultiplying (4.4) by λ^T and substituting (4.5), we get

$$\lambda^T A^T V^{-1} A \theta^* = \lambda^T A^T V^{-1} x. \tag{4.6}$$

Since (4.6) and (4.5) are identical with (4.1) and (4.2), the λ's and θ's must b the same, so the general Gaussian theorem is true.

This solution, which is similar to that of Rao, is satisfactory in proving th generalized Gaussian theorem but it is not satisfactory in that it does not provid an explicit value of θ^* (only implicit relations involving the vector paramete λ) nor does it give an explicit expression for the unbiased linear estimator havin minimum variance. These are provided in the sections following.

One further remark should be made before leaving these results on leas squares. The Eqs. (4.6) and (4.5) may be considered to be the normal equation of a general least squares problem expressed in terms of the vector parameter λ Comparison of (4.6) with (2.7) shows that these normal equations can be ob tained from the normal equations of the problem with no restrictions by pre multiplication by λ^T where λ^T is subject to the conditions $\lambda^T R^T = 0$.

5. The explicit form of the estimator. It appears that no one has provided th explicit form for ϕ^* or for θ^*. Post multiplication of (3.14) by $(A^T V^{-1} A)^{-1} R$

followed by application of [illegible] eliminates [illegible] with the resulting

[illegible]

Now since [illegible] is of order and rank [illegible], we can write

[illegible]

The value of [illegible] is then from [illegible]

[illegible]

and from [illegible]

[illegible]

so that

[illegible]

is the linear unbiased estimator having minimum variance, and

[illegible]

and [illegible] is the explicit solution of the normal equations. [illegible] did not give an explicit answer even for the case [illegible] since he did not derive an explicit formula for [illegible]. The argument above covers the [illegible] case with [illegible] and [illegible] vectors. Thus [illegible] and [illegible] hold with [illegible] a vector. As is pointed out above, the [illegible] which results from least squares and from minimum variance is independent of [illegible].

The results above are also general enough to include the Aitken results. These can be obtained formally from the above results by using the convention that [illegible] is [illegible] when [illegible], the formal equivalent of [illegible] side conditions. Thus the last terms drop from [illegible] and [illegible] for the Aitken problem.

The above results also generalize those of David and Neyman [illegible] who placed specifications on the dispersion matrix [illegible]. They defined [illegible] to be a diagonal matrix with

[illegible]

The formula for c^* then becomes

[illegible]

Now [illegible] is c^* with [illegible], and [illegible] is c^* with [illegible].

It follows that θ^* satisfies

$$\Lambda^T A^T V^{-1} A\theta^* = \Lambda^T A^T V^{-1} x. \tag{3.17}$$

Equations (3.17) and (3.12) may be considered to be basic relations in θ^* a
Λ^T.

4. The general Gaussian theorem. We next demonstrate the general Gaussi
theorem that the value of θ^* obtained by least squares is consistent with that
(3.17) and (3.12). We note first that θ^* in the general solution is an $s \times 1$ vect
and that the general solution is obtained by premultiplying θ^* by the fix
$k \times s$ matrix L. The general theorem is established by proving the typical ca
with $k = 1$ so that L, B, D, and Λ are vectors and (3.17) becomes

$$\lambda^T A^T V^{-1} A\theta^* = \lambda^T A^T V^{-1} x, \tag{4.1}$$

where λ^T is $\lambda^T(1 \times s)$. Also (3.12) becomes

$$\lambda^T R^T = 0. \tag{4.2}$$

Now we wish to minimize the scalar $Q = \epsilon^T V^{-1}\epsilon$, subject to the restriction co
ditions. Then

$$Q' = (A\theta - x)^T V^{-1}(A\theta - x) + 2(l - bA - dR)\lambda + 2\gamma R\theta. \tag{4.3}$$

Differentiation with respect to θ and d leads to the "normal" equations

$$A^T V^{-1} A\theta^* - A^T V^{-1} x + R^T \gamma^T = 0, \tag{4.4}$$

$$\lambda^T R^T = 0. \tag{4.5}$$

Premultiplying (4.4) by λ^T and substituting (4.5), we get

$$\lambda^T A^T V^{-1} A\theta^* = \lambda^T A^T V^{-1} x. \tag{4.6}$$

Since (4.6) and (4.5) are identical with (4.1) and (4.2), the λ's and θ's must b
the same, so the general Gaussian theorem is true.

This solution, which is similar to that of Rao, is satisfactory in proving th
generalized Gaussian theorem but it is not satisfactory in that it does not provid
an explicit value of θ^* (only implicit relations involving the vector paramet
λ) nor does it give an explicit expression for the unbiased linear estimator havin
minimum variance. These are provided in the sections following.

One further remark should be made before leaving these results on lea
squares. The Eqs. (4.6) and (4.5) may be considered to be the normal equatio
of a general least squares problem expressed in terms of the vector parameter
Comparison of (4.6) with (2.7) shows that these normal equations can be ob
tained from the normal equations of the problem with no restrictions by pr
multiplication by λ^T where λ^T is subject to the conditions $\lambda^T R^T = 0$.

5. The explicit form of the estimator. It appears that no one has provided th
explicit form for ϕ^* or for θ^*. Post multiplication of (3.14) by $(A^T V^{-1} A)^{-1}$

followed by application of (3.12) eliminates Λ^T with the resulting

(5.1) $$DR(A^TV^{-1}A)^{-1}R^T = L(A^TV^{-1}A)^{-1}R^T.$$

Now since $R(A^TV^{-1}A)^{-1}R^T$ is of order and rank u, we can write

(5.2) $$D = L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}.$$

The value of Λ^T is then from (3.14)

(5.2) $$\Lambda^T = L(A^TV^{-1}A)^{-1} - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}$$

and from (3.13),

(5.4) $$B = L(A^TV^{-1}A)^{-1}A^TV^{-1} - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}$$

so that

(5.5) $$\phi^* = L(A^TV^{-1}A)^{-1}A^TV^{-1}x - L(A^TV^{-1}A)^{-1}R^T [R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}x$$

is the linear unbiased estimator having minimum variance, and

(5.6) $$\theta^* = (A^TV^{-1}A)^{-1}A^TV^{-1}x - (A^TV^{-1}A)^{-1}R^T [R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}x,$$

and θ^* is the explicit solution of the normal equations. Rao did not give an explicit answer even for the case $k = 1$, since he did not derive an explicit formula for λ^T. The argument above covers the Rao case with L and λ vectors. Thus (5.5) and (5.6) hold with L a vector As is pointed out above, the θ^* which results from least squares and from minimum variance is independent of L.

The results above are also general enough to include the Aitken results. These can be obtained formally from the above results by using the convention that $R^T[R(A^TV^{-1}A)R^T]^{-1}$ is 0 when $R = 0$, the formal equivalent of $u = 0$ side conditions. Thus the last terms drop from (5.5) and (5.6) for the Aitken problem.

The above results also generalize those of David and Neyman [8] who placed specifications on the dispersion matrix V. They defined V to be a diagonal matrix with

(5.7) $$v_{ii} = \frac{\sigma^2}{P_{ii}}, \quad \text{where} \quad P_{ii} = \frac{\sigma^2}{\sigma_i^2}.$$

The formula for ϕ^* then becomes

(5.8) $$\phi^* = L(A^TPA)^{-1}A^TPx - L(A^TPA)^{-1}R^T[R(A^TPA)^{-1}R^T]^{-1}R(A^TPA)^{-1}A^TPx.$$

Now B is ϕ^* with $x = I$, and θ^* is ϕ^* with $L = I$.

It follows that θ^* satisfies

$$\Lambda^T A^T V^{-1} A \theta^* = \Lambda^T A^T V^{-1} x. \tag{3.17}$$

Equations (3.17) and (3.12) may be considered to be basic relations in θ^* a
Λ^T.

4. The general Gaussian theorem. We next demonstrate the general Gaussi
theorem that the value of θ^* obtained by least squares is consistent with that
(3.17) and (3.12). We note first that θ^* in the general solution is an $s \times 1$ vect
and that the general solution is obtained by premultiplying θ^* by the fix
$k \times s$ matrix L. The general theorem is established by proving the typical ca
with $k = 1$ so that L, B, D, and Λ are vectors and (3.17) becomes

$$\lambda^T A^T V^{-1} A \theta^* = \lambda^T A^T V^{-1} x, \tag{4.1}$$

where λ^T is $\lambda^T(1 \times s)$. Also (3.12) becomes

$$\lambda^T R^T = 0. \tag{4.2}$$

Now we wish to minimize the scalar $Q = \epsilon^T V^{-1} \epsilon$, subject to the restriction co
ditions. Then

$$Q' = (A\theta - x)^T V^{-1} (A\theta - x) + 2(l - bA - dR)\lambda + 2\gamma R\theta. \tag{4.3}$$

Differentiation with respect to θ and d leads to the "normal" equations

$$A^T V^{-1} A \theta^* - A^T V^{-1} x + R^T \gamma^T = 0, \tag{4.4}$$

$$\lambda^T R^T = 0. \tag{4.5}$$

Premultiplying (4.4) by λ^T and substituting (4.5), we get

$$\lambda^T A^T V^{-1} A \theta^* = \lambda^T A^T V^{-1} x. \tag{4.6}$$

Since (4.6) and (4.5) are identical with (4.1) and (4.2), the λ's and θ's must
the same, so the general Gaussian theorem is true.

This solution, which is similar to that of Rao, is satisfactory in proving t
generalized Gaussian theorem but it is not satisfactory in that it does not provi
an explicit value of θ^* (only implicit relations involving the vector paramet
λ) nor does it give an explicit expression for the unbiased linear estimator havi
minimum variance. These are provided in the sections following.

One further remark should be made before leaving these results on lea
squares The Eqs. (4.6) and (4.5) may be considered to be the normal equatio
of a general least squares problem expressed in terms of the vector parameter
Comparison of (4.6) with (2.7) shows that these normal equations can be o
tained from the normal equations of the problem with no restrictions by pr
multiplication by λ^T where λ^T is subject to the conditions $\lambda^T R^T = 0$.

5. The explicit form of the estimator. It appears that no one has provided t
explicit form for ϕ^* or for θ^*. Post multiplication of (3.14) by $(A^T V^{-1} A)^{-1} I$

followed by application of (3.12) eliminates Λ^T with the resulting

(5.1) $$DR(A^TV^{-1}A)^{-1}R^T = L(A^TV^{-1}A)^{-1}R^T.$$

Now since $R(A^TV^{-1}A)^{-1}R^T$ is of order and rank u, we can write

(5.2) $$D = L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}.$$

The value of Λ^T is then from (3.14)

(5.2) $$\Lambda^T = L(A^TV^{-1}A)^{-1} - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}$$

and from (3.13),

(5.4) $$B = L(A^TV^{-1}A)^{-1}A^TV^{-1} - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}$$

so that

(5.5) $$\phi^* = L(A^TV^{-1}A)^{-1}A^TV^{-1}x - L(A^TV^{-1}A)^{-1}R^T [R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}x$$

is the linear unbiased estimator having minimum variance, and

(5.6) $$\theta^* = (A^TV^{-1}A)^{-1}A^TV^{-1}x - (A^TV^{-1}A)^{-1}R^T [R(A^TV^{-1}A)^{-1}R^T]^{-1} R(A^TV^{-1}A)^{-1}A^TV^{-1}x,$$

and θ^* is the explicit solution of the normal equations. Rao did not give an explicit answer even for the case $k = 1$, since he did not derive an explicit formula for λ^T. The argument above covers the Rao case with L and λ vectors. Thus (5.5) and (5.6) hold with L a vector As is pointed out above, the θ^* which results from least squares and from minimum variance is independent of L.

The results above are also general enough to include the Aitken results. These can be obtained formally from the above results by using the convention that $R^T[R(A^TV^{-1}A)R^T]^{-1}$ is 0 when $R = 0$, the formal equivalent of $u = 0$ side conditions. Thus the last terms drop from (5.5) and (5.6) for the Aitken problem.

The above results also generalize those of David and Neyman [8] who placed specifications on the dispersion matrix V. They defined V to be a diagonal matrix with

(5.7) $$v_{ii} = \frac{\sigma^2}{P_{ii}}, \quad \text{where} \quad P_{ii} = \frac{\sigma^2}{\sigma_i^2}.$$

The formula for ϕ^* then becomes

(5.8) $$\phi^* = L(A^TPA)^{-1}A^TPx - L(A^TPA)^{-1}R^T[R(A^TPA)^{-1}R^T]^{-1}R(A^TPA)^{-1}A^TPx.$$

Now B is ϕ^* with $x = I$, and θ^* is ϕ^* with $L = I$.

If $P_{ii} = \sigma P'_{ii}$ with $P'_{ii} = 1/\sigma_i^2$, we have

(5.9)
$$\phi^* = L(A^TP'A)^{-1}A^TP'x - L(A^TP'A)^{-1}R^T[R(A^TP'A)^{-1}R^T]^{-1}R(A^TP'A)^{-1}A^TP'x.$$

Then dropping the side conditions on the parameters we get

(5.10)
$$B = L(A^TPA)^{-1}A^TP = L(A^TP'A)^{-1}A^TP'.$$

When L is restricted to a vector, this is the David-Neyman result in matrix form.

When $V = I$, $L = I$ and $R = 0$ we have the common case of unweighted least squares regression

$$\phi^* = \theta^* = (A^TA)^{-1}A^Tx$$

and

(5.11)
$$B = (A^TA)^{-1}A^T.$$

The general results are immediately applicable to a variety of special cases involving specifications on V, specifications on L, and specifications on R, separately or in combinations.

6. The dispersion matrix of solutions. The dispersion matrix of solutions is var $\phi^* = BVB^T$. Using the value of B in (5.4), we get

(6.1)
$$\text{var}(\phi^*) = BVB^T = L(A^TV^{-1}A)^{-1}L^T - L(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}R(A^TV^{-1}A)^{-1}L^T.$$

When $k = 1$, this is an explicit result for the Rao problem. When there are no side conditions we have the Aitken result

(6.2)
$$\text{var}(\phi^*) = L(A^TV^{-1}A)^{-1}L^T.$$

When the values of x are uncorrelated with $v_{ii} = \sigma^2/P_{ii}$, (6.1) and (6.2) become

(6.3)
$$\text{var}(\phi^*) = L(A^TPA)^{-1}L^T\sigma^2 - L(A^TPA)^{-1}R^T[R(A^TPA)^{-1}R^T]^{-1}R(A^TPA)^{-1}L^T\sigma^2$$

and

(6.4)
$$\text{var}(\phi^*) = L(A^TPA)^{-1}L^T\sigma^2.$$

When in addition the variables have a common variance σ^2, $\sigma_i^2 = \sigma^2$ and $P = I$. The Eqs. (6.3) and (6.4) appear with $(A^TA)^{-1}$ replacing $(A^TPA)^{-1}$.

If $\phi = \theta$, the above formulas appear with $L = I$. The simple case in which there are no side conditions, $\phi = \theta$, with variables uncorrelated but with equal variances gives

(6.5)
$$\text{var}(\theta^*) = (A^TA)^{-1}\sigma^2,$$

which is the formula for the dispersion matrix of regression coefficients in a common model.

7. Use of a matrix relation. The results (6.1) and (5.4) enable us to write a relation involving the value of B which gives the value of BVB^T having minimum diagonal terms and the resulting matrix. In order to write this relation in compact form we use

$$(7.1) \qquad C = (A^TV^{-1}A)^{-1} - (A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1}R(A^TV^{-1}A)^{-1},$$

which is Λ with $L = I$ to get

$$(7.2) \qquad BVB^T = LCL^T + (B - LCA^TV^{-1})V(B - LCA^TV^{-1})^T.$$

The relation used by Plackett ([1], page 459) is a special case of this relation with the terms involving R deleted. Then $C = (A^TV^{-1}A)^{-1}$. Plackett's relation may be considered to be a generalization of the relation used by Gauss in establishing the theorem. Once the relation is established we see at once that the diagonal terms of BVB^T are minimized for general B when

$$(7.3) \qquad B = LCA^TV^{-1}$$

as indicated in (5.4) and that the minimum values of the diagonal terms of the dispersion matrix are the diagonal terms of

$$(7.4) \qquad BVB^T = LCL^T$$

as given in (6.1).

Once this general relation (7.2) is proposed, it may be verified by direct expansion. Then the whole solution of the problem of the minimization of the diagonal terms of the dispersion matrix of the estimators is immediately available as indicated by Plackett. If the relation is not known, and it has not been known previously for the general problem, it can be established with the use of matrix derivatives as shown above.

The various special cases of the general matrix relation result from the application of specified conditions to V, L, and R.

8. The variance of the residuals. Returning to the problem of least squares, we call $\mathcal{E}(\epsilon^TV^{-1}\epsilon)$ the variance of the (weighted) residuals. Then ϵ can be written

$$(8.1) \qquad \epsilon = (ACA^TV^{-1} - I)z,$$

where C is given by (7.1), and ACA^TV^{-1}, and hence $ACA^TV^{-1} - I$, are idempotent. Hence

$$(8.2) \qquad \begin{aligned} \epsilon^TV^{-1}\epsilon &= z^T(ACA^TV^{-1} - I)^TV^{-1}(ACA^TV^{-1} - I)z \\ &= z^TV^{-1}z - z^TV^{-1}ACA^TV^{-1}z. \end{aligned}$$

There is no loss in generality, for purpose of derivation, in assuming that z [illegible] (8.1) and (8.2) is a deviate with [illegible] $= V$.

For the Aitken problem, $C = (A^TV^{-1}A)^{-1}$ and we have

$$(8.3) \qquad \epsilon^TV^{-1}\epsilon = z^TV^{-1}z - z^TV^{-1}A(A^TV^{-1}A)^{-1}A^TV^{-1}z.$$

To this we apply the triangular matrix square root transformation[1]

$$y = Wx \text{ with } W^TW = V^{-1}. \tag{8.4}$$

We then have

$$e^TV^{-1}e = y^Ty - y^TWA(A^TV^{-1}A)^{-1}A^TW^Ty \tag{8.5}$$

with

$$\text{var } (y) = E(Wxx^TW^T) = WVW^T = I_n \tag{8.6}$$

so that, using the trace

$$E(x^TV^{-1}x) = E(y^Ty) = n. \tag{8.7}$$

In order to find the expected value of the second term on the right in (8.5), w use the additional transformation

$$z = Sy \text{ with } S^TS = WA(A^TV^{-1}A)^{-1}A^TW^T, \tag{8.8}$$

where S is a triangular matrix. Since the rank of $WA(A^TV^{-1}A)^{-1}A^TW^T$ is s S is of rank s, and there are $n - s$ rows identically zero. Then

$$e^TV^{-1}e = y^Ty - z^Tz, \tag{8.9}$$

and since

$$E(zz^T) = \begin{bmatrix} I_s & 0 \\ 0 & 0 \end{bmatrix}, \tag{8.10}$$

then

$$E(z^Tz) = s$$

and

$$E(e^TV^{-1}e) = E(y^Ty) - E(z^Tz) = n - s. \tag{8.11}$$

In the general problem with more complex C we have the additional quadrati form

$$x^TV^{-1}A(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^TR^T]^{-1}R(A^TV^{-1}A)^{-1}\cdot A^TV^{-1}x \tag{8.12}$$

whose matrix is of rank u. Application of (8.4) followed by application of

$$t = U_y \text{, where } U^TU = WA(A^TV^{-1}A)^{-1}R^T[R(A^TV^{-1}A)^{-1}R^T]^{-1} \cdot R(A^TV^{-1}A)^{-1}A^TW^T \tag{8.13}$$

---

[1] A triangular matrix square root, as applied to this problem, is a triangular matrix W defined by $W^TW = V^{-1}$. This should not be confused with the (non-triangular) algebrai matrix square root defined by $(V^{-\frac{1}{2}})^2 = V^{-1}$.

reduces this term to

(8.14) $$t^T t \text{ with } E(t^T t) = u.$$

Then

(8.15) $$E(e^T V^{-1} e) = n - s + u = n - (s - u).$$

This result is what one would expect. If the values of x were distributed normally, the positive definite quadratic form $\epsilon^T V^{-1} \epsilon$ would be distributed as χ^2 with $E(X^2) = n - (s - u)$ indicates the number of *independent* parameters.

This result is independent of k. In the Rao problem, $k = 1$, and the value of $E(\epsilon^T V^{-1} \epsilon)$ is $n - s + u$ as above. For the Aitken problem, $u = 0$, and the value is $n - s$. Where $V^{-1} = P/\sigma^2$ we have

(8.16) $$E(\epsilon^T P \epsilon) = (n - s + u)\sigma^2$$

and when $u = 0$, this is

(8.17) $$E(\epsilon^T P \epsilon) = (n - s)\sigma^2$$

as shown by David and Neyman for the case of uncorrelated variables ([8], pages 110–112). When $P = I$ this becomes

(8.18) $$E(\epsilon^T \epsilon) = (n - s)\sigma^2$$

as shown by Aitken using the properties of idempotent matrices ([3], page 139).

9. An estimator of the dispersion matrix of ϕ^*. David and Neyman [8] have provided an unbiased estimate of var ϕ^* for the case in which $V^{-1} = P/\sigma^2$, the x's are uncorrelated and L is a vector. A generalization related to the David-Neyman formula for the general problem is, for known V,

(9.1) $$E^{-1} \operatorname{var}(\phi^*) = \frac{\epsilon^T V^{-1} \epsilon}{n - s + u} LCL^T,$$

since its expected value is the dispersion matrix of ϕ^*.

When V is known this formula is of little value since BVB^T can be computed and no estimation is necessary. However if V is not known, but P is, we have

(9.2) $$E^{-1} \operatorname{var}(\phi^*) = \frac{\epsilon^T P \epsilon}{n - s + u} \cdot L\{(A^T PA)^{-1} - (A^T PA)^{-1} R^T [R(A^T PA)^{-1} R^T]^{-1} R(A^T PA)^{-1}\} L^T.$$

When $P = I$, the case of equal variances, we have the important

(9.3) $$E^{-1} \operatorname{var}(\phi^*) = \frac{\epsilon^T \epsilon}{n - s + u} L\{(A^T A)^{-1} - (A^T A)^{-1} R^T [R(A^T A)^{-1} R^T]^{-1} R(A^T A)^{-1}\} L^T.$$

In the case of no side conditions we have

(9.4) $$E^{-1} \text{ var } (\phi^*) = \frac{\epsilon^T P \epsilon}{n - s} L(A^T P A)^{-1} L^T.$$

Using the value of B in (5.10) we get

(9.5) $$E^{-1} \text{ var } (\phi^*) = \frac{\epsilon^T P \epsilon}{n - s} B P^{-1} B^T.$$

If L is a vector, the estimate is a scalar. In the David-Neyman scalar notation, with the x's uncorrelated and B a row vector (λ) we have

(9.6) $$\mu_1^2 = \frac{S_0}{n - s} \sum_{i=1}^{n} \frac{\lambda_i^2}{P_i},$$

where $[\lambda_i] = \lambda = L(A^T P A)^{-1} A^T P = B$. Hence (9.2) gives the estimator matrix of var (ϕ^*) for a more general problem than does (9.6).

Appendix Showing Orders of Matrices and Conditions

Matrix	Order	Matrix	Order
X	$n \times 1$	ψ	$k \times k$
A	$n \times s$	Λ	$s \times k$
θ and θ^*	$s \times 1$	λ	$s \times 1$
ϵ	$n \times 1$	R	$u \times s$
V	$n \times n$	g	$u \times 1$
Q and Q'	1×1	D	$k \times u$
$A^T V^{-1} A$	$s \times s$	μ	$k \times 1$
L	$k \times s$	γ	$1 \times u$
ϕ and ϕ^*	$k \times 1$	$R(A^T V^{-1} A)^{-1} R^T$	$u \times u$
B	$k \times n$	C	$s \times s$
var ϕ^*	$k \times k$	P	$n \times n$
BVB^T	$k \times k$		

$u < s < n$, $u = 0$ gives Aitken problem, $k = 1$ gives Rao problem, $V^{-1} = P/\sigma^2$ gives David-Neyman condition.

REFERENCES

[1] R. L. PLACKETT, "A historical note on the method of least squares," *Biometrika*, Vol. 36 (1949), pp. 458–460.

[2] A. C. AITKEN, "On least squares and linear combinations of observations," *Proc. Roy. Soc. Edin.*, Vol. 55 (1934), pp. 42–48.

[3] A. C. AITKEN, "Studies in mathematics. IV. On linear approximations by least squares," *Proc. Roy. Soc. Edin.*, Vol. 62 (1945), pp. 138–146.

[4] C. R. RAO, "Generalization of Markoff's theorem and tests of linear hypotheses," *Sankhya*, Vol. 7 (1945), pp. 9–16.

[5] C. R. RAO, "Markoff's Theorem with linear restrictions on the parameters," *Sankhya*, Vol. 7 (1945), pp. 16–19.
[6] P. S. DWYER, "A matrix presentation of least squares and correlation theory," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 82–89.
[7] P. S. DWYER AND M. S. MACPHAIL, "Symbolic matrix derivatives," *Ann Math. Stat.*, Vol. 19 (1948), pp. 517–534.
[8] F. N. DAVID AND J. NEYMAN, "Extension of the Markoff Theorem on least squares," *Statistical Research Memoirs*, Vol. 2 (1938), pp. 105–116.

The following simple lemmas which we present without proof are useful. Let X and Y designate the vectors $(X_1, X_2, \cdots, X_k)$ and $(Y_1, Y_2, \cdots, Y_k)$.

LEMMA 1. *If* $X_1, X_2, \cdots, X_k$ *are interchangeable* r.v.'s *and* $Y = \psi(X)$ *is defined by* $Y_j = \Phi[X_j, g(X)], j = 1, 2, \cdots, k$, *where* Φ *and* g *are Borel measurable functions, the latter being symmetric in its* k *arguments, then* $Y_1, Y_2, \cdots, Y_k$ *are interchangeable.*

LEMMA 2. *If* $Y = (Y_1, Y_2, \cdots, Y_k)$ *is a random permutation of the interchangeable* r.v.'s $X_1, X_2, \cdots, X_k$, *then* Y *has the same distribution as* X.

3. Background and Framework. The term "Central Limit Theorem" is a loose designation for one of an agglomeration of theorems dealing with limiting normality of distributions of sums of random variables—in the classical treatment—independent random variables.

The early results of De Moivre and Laplace have been succeeded by ever more powerful theorems set in an increasingly general framework. Recent works [5], [6] commence with a double sequence of rowwise *independent* r.v.'s (i.e., the r.v.'s within each row are independent)

$$\begin{array}{cccc} X_{11}, & X_{12}, & \cdots, & X_{1k_1} \\ X_{21}, & X_{22}, & \cdots, & X_{2k_2} \\ \vdots & \vdots & & \\ X_{n1}, & X_{n2}, & \cdots, & X_{nk_n} \\ \vdots & \vdots & \cdots & \vdots \end{array}$$

(where $k_n \to \infty$) and investigate the limiting distributions, i.e., c.d.f.'s of the row sums, say $S_n = \sum_{k=1}^{k_n} X_{nk}$. To render the problem more meaningful the r.v.'s are required to be "infinitesimal" (or asymptotically constant), i.e.,

$$\lim_{n\to\infty} \max_{1\leq i\leq k_n} P\{|X_{ni}| > \epsilon\} = 0, \qquad \text{all} \quad \epsilon > 0.$$

A famous theorem of Khintchine asserts that the class of limiting distribution of such sums S_n coincides with the class of infinitely divisible laws [5]. A necessary and sufficient condition that the limiting distribution (assuming one exists) of sums of row-wise independent infinitesimal r.v.'s be normal is well known, namely, $\max_{1<i<k_n} |X_{ni}| \xrightarrow{P} 0$. (This actually implies infinitesimality here). For purposes of comparison with Theorem 1 of the next section we state the following result of Raikov (cf. [5]):

If Z_{nk}, $k = 1, \cdots, k_n$ are infinitesimal rowwise independent r.v.'s with zero means and finite variances σ_{nk}^2 with $\sum_{k=1}^{k_n} \sigma_{nk}^2 = 1$, a necessary and sufficient condition that the c.d.f. of $\sum_{k=1}^{k_n} Z_{nk}$ converges to the normal c.d.f. with mean 0 and variance 1 is that $\sum_{k=1}^{k_n} Z_{nk}^2 \xrightarrow{P} 1$.

Attempts have been made to relax the requirement of independence with varying degrees of success. Perhaps a natural and useful generalization is to double sequences of interchangeable random variables.

In this direction, let X_{ni}, $i = 1, \cdots, k_n$ comprise a (finite) set of i.r.v.'s for every $n = 1, 2, \cdots,$.

If we stipulate that $\lim_{n\to\infty} P\{|X_{n1}| > \epsilon\} = 0$, all $\epsilon > 0$, the question of the nature of the class C^* of all limiting distributions of row sums may again be posed. Clearly, C^* includes all stable distributions but contains others as well. This follows from a result of von Mises [7] who showed that the distribution of the number S_{n,r_n} of unoccupied cells in a random casting of r_n objects into n cells approaches that of the Poisson when n, $r_n \to \infty$ in a manner such that the expected number of vacancies is constant. If the expected proportion of vacancies converges to a constant, then Irving Weiss [9] has shown that the limiting distribution, suitably normalized, is normal. But $S_{n,r_n} = \sum_{i=1}^{n} X_{ni}$ where the X_{ni} are i.r.v.'s assuming the values one or zero (according as the ith cell is empty or not). Therefore, the Poisson distribution and in fact all infinitely divisible distributions belong to C^*.

In this paper, we consider only the case of limiting normal distributions and treat the first of the following two situations:

(a) For each $n = 1, 2, \cdots$, the i.r.v.'s X_{ni}, $i = 1, 2, \cdots, k_n$ have a non-random sum.

(b) For each $n = 1, 2, \cdots$, the i.r.v.'s X_{ni}, $i = 1, \cdots, k_n$ are embeddable in an infinite sequence of i.r.v.'s.[2]

These cases are mutually exclusive since if $\sum_{i=1}^{k_n} X_{ni} = C_n$, the covariance of any pair of i.r.v.'s equals $-[1/(k_n - 1)]$ multiplied by the common variance. But then their correlation is negative, which is precluded (under case b) by a prior remark.

4. I.R.V.'s whose sum is non-random. For each $n = 1, 2, \cdots$, let X'_{nk}, $k = 1, 2, \cdots, k_n(\to \infty)$ be i.r.v.'s with finite variance $\sigma'^2_{nk} = \sigma'^2_{n1} = E(X'_{n1} - EX'_{n1})^2$ and satisfying the linear constraint

$$\text{(i')} \qquad \sum_{i=1}^{k_n} X'_{ni} = C_n .$$

Naturally, under such a proviso we must investigate partial rather than complete row sums.

If we define

$$X_{ni} = \frac{1}{\sigma'_{ni}}\left(X'_{ni} - \frac{C_n}{k_n}\right),$$

the X_{ni} are, by Lemma 1, i.r.v.'s satisfying the relationships

$$\text{(i)} \qquad \sum_{i=1}^{k_n} X_{ni} = 0, \qquad n = 1, 2, \cdots,$$

and

(ii) $EX^2_{ni} = \sigma^2_{X_{ni}} = 1, i = 1, 2, \cdots, k_n$ and all $n = 1, 2, \cdots$,

[2] The theorems obtained for the case of infinite sequences of i.r.v.'s overlap results of Professors Blum and Rosenblatt of Indiana University and will appear in a joint publication.

We suppose, therefore, without loss of generality, that for each

$$n = 1, 2, \cdots, \{X_{nk}\}, k = 1, \cdots, k_n(\to \infty)$$

are rowwise i.r.v.'s satisfying (i) and (ii) and possessing the joint $F_n(x_1, x_2, \cdots, x_{k_n})$. We have then

THEOREM 1. *For each* $n = 1, 2, \cdots$, *let* $\{X_{ni}\}$ *be interchangeable random ables satisfying* (i) *and* (ii). *If*

$$\max_{1 \leq k \leq k_n} \frac{|X_{nk}|}{\sqrt{k_n}} \xrightarrow{P} 0, \tag{1}$$

$$\frac{1}{k_n} \sum_{k=1}^{k_n} X_{nk}^2 \xrightarrow{P} 1, \tag{2}$$

and $m_n < k_n$ *is a sequence of positive integers such that* $\lim_{n\to\infty} m_n/k_n = \alpha$, $\alpha < 1$, *then*

$$\lim_{n\to\infty} P\left\{\frac{1}{\sqrt{m_n}} \sum_{k=1}^{m_n} X_{nk} < x\right\} = \frac{1}{\sqrt{2\pi(1-\alpha)}} \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2(1-\alpha)}\right) dy$$

PROOF. For any set of real numbers x_{ni}, if $\max_{1\leq i\leq k_n} |x_{ni}|/\sqrt{k_n} = o(1)$ $\lim_{n\to\infty} (1/k_n) \sum_1^{k_n} x_{ni}^2 = 1$, then

$$\max_{1 \leq i \leq k_n} \frac{|x_{ni}|}{\sqrt{\sum_1^{k_n} x_{ni}^2}} = o(1).$$

It follows directly that if the x_{ni} are r.v.'s and the analogous conditions are in "probability" the conclusion holds "in probability." That is, (1) and imply

$$\max_{1 \leq i \leq k_n} \frac{|X_{ni}|}{\sqrt{\sum_{i=1}^{k_n} X_{ni}^2}} \xrightarrow{P} 0. \tag{5}$$

Next, let $Y_{n1}, \cdots, Y_{nk_n}$ be a randomly selected permutation of $X_{n1}, \cdots,$. Then even when it is stipulated that X_{ni} = fixed real number $i = 1, 2, \cdots, k_n$, the quantity

$$U_n = \left(\sum_{i=1}^{k_n} X_{ni}^2\right)^{-1/2} \sum_{i=1}^{m_n} Y_{ni}$$

is a random variable.

Suppose that for some c.d.f. $G(u)$ and arbitrary $\epsilon > 0$, there exists δ_ϵ and integral $N_1(\epsilon)$ (all independent of $x_{n1}, \cdots, x_{nk_n}$) such that

$$\max_{1 \leq i \leq k_n} \frac{|x_{ni}|}{\sqrt{\sum_1^{k_n} x_{ni}^2}} < \delta_\epsilon$$

implies

$$|P\{U_n < u | X_{ni} = x_{ni}, i = 1, \cdots, k_n\} - G(u)| < \epsilon \tag{6}$$

for all $n > N_1(\epsilon)$ and continuity points u of $G(u)$. By (5), there exists $N_2(\epsilon)$ such that for all $n > N_2(\epsilon)$, say

$$\epsilon > P\left\{\max_{1 \leq i \leq k_n} \frac{|X_{ni}|}{\sqrt{\sum_1^{k_n} X_{ni}^2}} > \delta_\epsilon\right\} = P\{\bar{A}_n\}. \tag{7}$$

Then, from (6) and (7) for arbitrary $\epsilon > 0$ and $n > \max[N_1(\epsilon), N_2(\epsilon)]$ and continuity points u of $G(u)$,

$$\begin{aligned} &| P\{U_n < u\} - G(u) | \\ &= \left| \int_{R^{k_n}} [P\{U_n < u \mid X_{ni} = x_{ni}, i = 1, \cdots, k_n\} - G(u)] \, dF_n(x_1, \cdots, x_{k_n}) \right| \qquad (8) \\ &\leqq \int_{A_n} | P\{U_n < u \mid X_{ni}, i = 1, \cdots, k_n\} - G(u) | \, dF_n + \int_{\bar{A}_n} dF_n \leqq 2\epsilon. \end{aligned}$$

For simplicity in writing, let Q be an r.v with c.d.f. $G(u)$; then for $\lambda > 0$, $Q_\lambda = (1/\lambda)\, Q$ is an r.v. with distribution $G(\lambda u)$. Under the proviso (6), (8) shows that

$$U_n = \frac{\sum_{i=1}^{m_n} Y_{ni}}{\sqrt{\sum_1^{k_n} X_{ni}^2}} \xrightarrow{L} Q.$$

On the other hand, according to (2),

$$\sqrt{\frac{1}{k_n} \sum_{i=1}^{k_n} X_{ni}^2} \xrightarrow{P} 1.$$

Consequently, (see, e.g., [3]),

$$\frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} Y_{ni} = \sqrt{\frac{k_n}{m_n}}\, U_n \sqrt{\frac{1}{k_n} \sum_{i=1}^{k_n} X_{ni}^2} \xrightarrow{L} \frac{1}{\sqrt{\alpha}} Q = Q_{\sqrt{\alpha}}.$$

But by Lemma 2, $\sum_{i=1}^{m_n} Y_{ni}$ and $\sum_{i=1}^{m_n} X_{ni}$ have the same distribution, and thus, under the proviso (6),

$$\frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} X_{ni} \xrightarrow{L} Q_{\sqrt{\alpha}}\,. \tag{9}$$

It remains to verify (6) for $G(u)$ the c.d.f. of $Q = N_{0,\alpha(1-\alpha)}$, where N_{μ,σ^2} represents a normal random variable with mean μ and variance σ^2. To do so, it suffices to prove that

$$U_n = \left(\sum_{i=1}^{k_n} x_{ni}^2\right)^{-1/2} \sum_{i=1}^{m_n} Y_{ni} \xrightarrow{L} Q = N_{0,\alpha(1-\alpha)}\,, \tag{10}$$

providing $Y_{n1}, Y_{n2}, \cdots, Y_{nk_n}$ is a random permutation of the fixed

$x_{n1}, x_{n2}, \cdots, x_{nk_n}$, where

(11)
$$\max_{1 \leqq k \leqq k_n} \left(\sum_{i=1}^{k_n} x_{ni}^2 \right)^{-1/2} |x_{nk}| = o(1)$$

and $\sum_{i=1}^{k_n} x_{ni} = 0$. A theorem of Noether [8] states that the distribution of

$$L_n = \sum_{i=1}^{k_n} d_{ni} Y_{ni}$$

converges to the normal distribution (when normalized by its mean and stand deviation) if the d_{ni} are fixed real numbers such that

$$D_{n,s} = \frac{\frac{1}{k_n} \sum_{i=1}^{k_n} (d_{ni} - \bar{d}_n)^s}{\left[\frac{1}{k_n} \sum_{i=1}^{k_n} (d_{ni} - \bar{d}_n)^2 \right]^{s/2}} = O(1) \qquad \text{for } s = 3, 4,$$

and

$$A_{ns} = \frac{\sum_{i=1}^{k_n} (x_{ni} - \bar{x}_n)^s}{\left[\sum_{i=1}^{k_n} (x_{ni} - \bar{x}_{ni})^2 \right]^{s/2}} = o(1), \qquad \text{for } s = 3, 4,$$

with $\bar{d}_n = 1/k_n \sum_{i=1}^{k_n} d_{ni}$ and $\bar{x}_r = 1/k_n \sum_{i=1}^{k_n} x_{ni} = 0$. Let $d_{ni} = 1$ for $1 \leqq i \leqq$ and 0 for $m_n + 1 \leqq i \leqq k_n$. Then $D_{n,s} = 0(1)$ for $s = 3, 4, \cdots$. Furtherm from (11),

$$A_{ns} = \frac{\sum_{i=1}^{k_n} x_{ni}^s}{\left(\sum_{i=1}^{k_n} x_{ni}^2 \right)^{s/2}} \leqq \frac{\max_{1 \leqq i \leqq k_n} |x_{ni}|^{s-2}}{\left(\sum_{i=1}^{k_n} x_{ni}^2 \right)^{s-2/2}} = o(1) \quad \text{for } s = 3, 4, \cdot$$

Thus Noether's theorem applies to $L_n = \sum_{i=1}^{m_n} Y_{ni}$ whose mean and varia we shall show to be given by $\mu_n = 0$ and

(12)
$$\sigma_n^2 = \left(\sum_{i=1}^{k_n} x_{ni}^2 \right) \frac{m_n(k_n - m_n)}{k_n(k_n - 1)}.$$

Then we shall have $L_n/\sigma_n \xrightarrow{L} N_{0,1}$ and the desired result

$$U_n = \frac{L_n}{\sigma_n} \sqrt{\frac{m_n(k_n - m_n)}{k_n(k_n - 1)}} \xrightarrow{L} N_{0,\alpha(1-\alpha)}.$$

We now conclude by evaluating μ_n and σ_n.

$$E(Y_{ni}) = \sum_{a=1}^{k_n} x_{na}/k_n = 0, \qquad E(Y_{ni}^2) = \sum_{a=1}^{n} x_{na}^2/k_n,$$

$$E(Y_{ni} Y_{nj}) = (\sum_{a \neq b} x_{na} x_{nb})/k_n(k_n - 1) = - \sum_{a=1}^{n} x_{na}^2/k_n(k_n - 1), \qquad i \neq$$

Hence

$$\mu_n = E\left(\sum_{i=1}^{m_n} Y_{ni}\right) = 0$$

and

$$\sigma_n^2 = E\left(\sum_{i=1}^{m_n} Y_{ni}\right)^2 = \left(\sum_{i=1}^{k_n} x_{ni}^2\right)\left(\frac{m_n}{k_n} - \frac{m_n(m_n-1)}{k_n(k_n-1)}\right),$$

which matches (12), concluding the proof.

COROLLARY 1. *For each positive integer* $n = 1, 2, \cdots$, *let* $\{X_{ni}\}$, $i = 1, 2, \cdots, k_n(\to \infty)$ *be i.r.v.'s satisfying* (i) *and* (ii). *If* $m_n < k_n$ *is a sequence of positive integers with* $\lim_{n\to\infty} m_n/k_n = \alpha$, $0 < \alpha < 1$ *and*

$$E[X_{n1}^4] = o(k_n), \qquad \mathrm{Cov}(X_{n1}^2, X_{n2}^2) = o(1), \tag{3}$$

then the conclusion of the theorem holds.

PROOF. For any $\eta > 0$,

$$P\left\{\max_{1\le i\le k_n} \frac{|X_{ni}|}{\sqrt{k_n}} > \eta\right\} = P\left\{\bigcup_1^{k_n}\left[\frac{|X_{ni}|}{\sqrt{k_n}} > \eta\right]\right\}$$

$$\le k_n P\left\{\frac{|X_{n1}|}{\sqrt{k_n}} > \eta\right\} \le \frac{k_n E|X_{n1}|^4}{k_n^2\eta^2} = o(1)$$

and

$$P\left\{\left|\frac{1}{k_n}\sum_1^{k_n}(X_{ni}^2 - 1)\right| > \eta\right\} \le \frac{E\left[\sum_1^{k_n}(X_{ni}^2 - 1)\right]^2}{k_n^2\eta^2}$$

$$= \frac{k_n E(X_{n1}^2 - 1)^2 + k_n(k_n - 1)\,\mathrm{Cov}\,(X_{n1}^2, X_{n2}^2)}{k_n^2\eta^2} = o(1).$$

COROLLARY 2. *For each* $n = 1, 2, \cdots$, *let* $\{X'_{ni}\}$, $i = 1, 2, \cdots, k_n(\to \infty)$ *be i.r.v.'s with* $\sum_{i=1}^{k_n} X'_{ni} = C_n$ *and* $\sum_{i=1}^{k_n}(X'_{ni})^2 = D_n^2 > 0$. *If*

$$\max_{1\le i\le k_n} \frac{|X'_{ni} - C_n/k_n|}{\sqrt{1/k_n(D_n^2 - C_n^2/k_n)}} \xrightarrow{P} 0,$$

then the conclusion of the theorem holds for $(1/\sqrt{m_n})\sum_1^{m_n} X_{ni}$, *where*

$$X_{ni} = \frac{X'_{ni} - C_n/k_n}{[(1/k_n)(D_n^2 - C_n^2/k_n)]^{\frac{1}{2}}}.$$

PROOF. Condition (2) is certainly satisfied since $1/k_n\sum_1^{k_n} X_{ni}^2 = 1$.

COROLLARY 3. *For each* $n = 1, 2, \cdots$, *let* $\{X_{ni}\}$, $i = 1, \cdots, k_n(\to \infty)$ *be i.r.v.'s with* $EX_{n1} = 0$, $EX_{n1}^2 = 1$ *and* $\bar{X}_n = 1/k_n\sum_1^{k_n} X_{nk}$. *If the* $\{X_{ni}\}$ *satisfy* (1), (2), *and*

$$E(X_{n1}X_{n2}) = o(1), \tag{4}$$

then

$$\lim_{n\to\infty} P\left\{\frac{1}{\sqrt{m_n}}\sum_1^{m_n}(X_{ni}-\bar{X}_n) < x\right\} = \frac{1}{\sqrt{2\pi(1-\alpha)}}\int_{-\infty}^{x}\exp\left(-\frac{y^2}{2(1-\alpha)}\right)dy.$$

PROOF. Let

$$Y_{ni} = \sqrt{\frac{k_n}{k_n-1}}\frac{X_{ni}-\bar{X}_n}{\sqrt{1-EX_{n1}X_{n2}}} = a_n(X_{ni}-\bar{X}_n).$$

Then, applying Lemma 1 with $g(X) = \bar{X}$, it follows that the $\{Y_{ni}\}$ are i.r.v.'s. Further, $\sum_{i=1}^{k_n} Y_{ni} = 0$ and $EY_{ni}^2 = 1$. Since

$$0 \leqq \max_{1\leqq i\leqq k_n}\frac{|Y_{ni}|}{\sqrt{k_n}} \leqq \frac{2a_n}{\sqrt{k_n}}\max_{1\leqq i\leqq k_n}|X_{ni}|,$$

(1) and (4) imply

$$\max_{1\leqq i\leqq k_n}\frac{|Y_{ni}|}{\sqrt{k_n}} \xrightarrow{P} 0.$$

Next, for every $\epsilon > 0$,

$$P\left\{\frac{1}{k_n}\left|\sum_{i=1}^{k_n}X_{ni}\right| > \epsilon\right\} \leqq \frac{E\left(\sum_{i=1}^{k_n}X_{ni}\right)^2}{k_n^2\,\epsilon^2} = (k_n\epsilon^2)^{-1} + \frac{(k_n-1)}{k_n}\frac{EX_{n1}X_{n2}}{\epsilon^2} = o(1).$$

That is, $\bar{X}_n \xrightarrow{P} 0$. Thus,

$$\frac{1}{k_n}\sum_{i=1}^{k_n}Y_{ni}^2 = \frac{(k_n-1)^{-1}}{(1-EX_{n1}X_{n2})}\left[\sum_{j=1}^{k_n}X_{nj}^2 - k_n\bar{X}_n^2\right] \xrightarrow{P} 1.$$

A direct application of the theorem to the $\{Y_{ni}\}$ shows that

$$\sqrt{\frac{k_n}{(k_n-1)}\frac{1}{(1-EX_{n1}X_{n2})}}\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}(X_{ni}-\bar{X}_n) \xrightarrow{L} N_{0,1-\alpha},$$

which, in view of (4), implies that

$$\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}(X_{ni}-\bar{X}_n) \xrightarrow{L} N_{0,1-\alpha}.$$

COROLLARY 4. *For each* $n = 1, 2, \cdots$, *let* $\{X_{ni}\}$, $i = 1, \cdots, k_n$, *be i.r.v.'s with* $EX_{n1} = 0$, $EX_{n1}^2 = 1$. *If* m_n *is a sequence of positive integers such that* $\lim_n m_n/k_n = \alpha$, $0 < \alpha < 1$, *and the* $\{X_{ni}\}$ *satisfy*

$$\text{Cov}(X_{n1}, X_{n2}) = \frac{-1}{k_n-1} + o(k_n^{-1}) \tag{4'}$$

and either (3) *or* (1) *and* (2), *then*

$$\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}X_{ni} \xrightarrow{L} N_{0,1-\alpha}.$$

PROOF. Since, as shown in the proof of Corollary 1, (3) implies (1) and (2), it suffices to suppose that the latter obtain. But (4′) clearly implies (4) whence, according to Corollary 3,

$$\frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} X_{ni} - \sqrt{m_n}\bar{X}_n \xrightarrow{L} N_{0,1-\alpha}.$$

However, for positive arbitrary positive,

$$\begin{aligned} P\{|\sqrt{m_n}\bar{X}_n| > \epsilon\} &\leqq \frac{m_n}{\epsilon^2 k_n^2} E\left(\sum_{i=1}^{k_n} X_{ni}\right)^2 \\ &= \frac{m_n}{\epsilon^2 k_n}\left[1 + (k_n - 1)\left\{\frac{-1}{k_n - 1} + o(k_n^{-1})\right\}\right] \\ &= \frac{m_n}{\epsilon^2 k_n} o(1) \end{aligned}$$

employing (4′). Thus, $\sqrt{m_n}\,\bar{X}_n \xrightarrow{P} 0$ and $1/\sqrt{m_n}\sum_{i=1}^{m_n} X_{ni} \xrightarrow{L} N_{0,1-\alpha}$.

In this instance, not only does $\bar{X}_n \xrightarrow{P} 0$, but even $1/\sqrt{k_n}\sum_{i=1}^{k_n} X_{ni} \xrightarrow{P} 0$, which is perhaps more than might be desired. Note that (4′) automatically prevails if the X_{ni} sum to C_n; in fact, $\text{Cov}(X_{n1}, X_{n2}) = -[1/(k_n - 1)]$ in this case.

Define $Z_{ni} = X_{ni}/\sqrt{k_n}$. If (i) is replaced by (iii), $EX_{ni} = 0$, and (ii) still obtains, then $EZ_{ni} = 0$, $\sum_{i=1}^{k_n} \sigma_{Z_{ni}}^2 = 1$. Conditions (1) and (2) become

(1′) $$\max_{1 \leqq i \leqq k_n} |Z_{ni}| \xrightarrow{P} 0$$

and

(2′) $$\sum_{i=1}^{k_n} Z_{ni}^2 \xrightarrow{P} 1.$$

Then, in view of theorems cited in Section 3, the conditions (2′) implies (1′) (and correspondingly (2) implies (1)) for infinitesimal row-wise *independent* r.v.'s, satisfying (ii) and (iii).

Of course, condition (i) precludes independence. Nonetheless, it should be verified for interchangeable r.v.'s satisfying (i) and (ii) that conditions (1) and (2) do not overlap. This may be seen from the following examples:

EXAMPLE 1. Let $(X_{n1}, X_{n2}, \cdots, X_{n,2n})$ be a random permutation of

$$(\sqrt{n}, -\sqrt{n}, 0, 0, \cdots, 0).$$

Then $\sum_{i=1}^{2n} X_{ni} = 0$, $1/2n \sum_{i=1}^{2n} X_{ni}^2 = 1$, but $\max_{1 \leqq i \leqq 2n} |X_{ni}|/\sqrt{2n} = 1/\sqrt{2}$.

EXAMPLE 2. Let $X = (X_{n1}, X_{n2}, \cdots, X_{n,2n}) = (0, 0, \cdots, 0)$ with probability $1 - p_n \to 1$, and otherwise let X be a random permutation of

$$\left(\frac{1}{\sqrt{p_n}}, -\frac{1}{\sqrt{p_n}}, \frac{1}{\sqrt{p_n}}, \cdots, -\frac{1}{\sqrt{p_n}}\right).$$

Then $\sum_{i=1}^{2n} X_{ni} = 0$, $E(X_{ni}^2) = 1$, and the X_{ni} are i.r.v.'s. Now

$$\max_{1 \leq i \leq 2n} \frac{|X_{ni}|}{\sqrt{2n}} = \frac{|X_{n1}|}{\sqrt{2n}} \xrightarrow{P} 0.$$

But $1/2n \sum_{i=1}^{2n} X_{ni}^2 = 0$ with probability $1 - p_n \to 1$ and hence converges t
zero in probability.

5. Illustrations.

EXAMPLE 1. *Quantiles.* Let k, n be positive integers and $U_1, U_2, \cdots, U_{kn}$
independent r.v.'s each uniformly distributed on (0, 1). Take U_j^* to be the jt
smallest of $(U_1, U_2, \cdots, U_{kn-1})$, $j = 1, 2, \cdots, kn - 1$. That i
$U_1^* \leqq U_2^* \leqq, \cdots, \leqq U_{kn-1}^*$ are the order statistics from a uniform or rectangu
lar distribution. Designate the successive differences $U_i^* - U_{i-1}^*$ by V_i, i =
$1, 2, \cdots, kn$, where $U_0^* = 0$, $U_{kn}^* = 1$.

It is well known that $V_1, V_2, \cdots, V_{kn}$ are interchangeable random variabl
adding up to one. In fact, any $kn - 1$ of them have a joint density

$$f(v_1, v_2, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{kn}) = (kn - 1)! \quad \text{for} \quad \textstyle\sum_{j \neq i} v_j \leqq 1, v_j \geqq 0,$$
$$= 0, \qquad \text{otherwise.}$$

A routine but tedious calculation or a non-routine exciting application of th
Poisson stochastic process yields

$$E[V_1^r] = \binom{kn - 1 + r}{r}^{-1}, \qquad r = 1, 2, \cdots$$

$$E[V_1^2 V_2^2] = \frac{4(kn - 1)!}{(kn + 3)!},$$

$$E[V_1 V_2] = \frac{(kn - 1)!}{(kn + 1)!},$$

$$E[V_1^2 V_2] = \frac{2(kn - 1)!}{(kn + 2)!}.$$

Further, $V_1, \cdots, V_{kn}$ are i.r.v.'s and likewise $X_{n1}, \cdots, X_{n,k_n}$, where $k_n = k$
and

$$X_{ni} = \frac{kn\left(V_i - \frac{1}{nk}\right)}{\sqrt{(kn - 1)(kn + 1)^{-1}}}, \qquad i = 1, 2, \cdots, k_n$$

Moreover, $\sum_{i=1}^{k_n} X_{ni} = 0$ and $\sigma_{X_{ni}}^2 = 1$, $i = 1, \cdots, kn$. The prior array of ex
pected values furnishes the estimates:

$$EX_{n1}^4 = O(n^4)E\left[V_1 - \frac{1}{kn}\right]^4 = O(n^4)O\left(\frac{1}{n^4}\right) = O(1)$$

and

$$\text{Cov}\,(X_{n1}^2, X_{n2}^2)$$

$$= O(n^4)\,\text{Cov}\left[\left(V_1 - \frac{1}{kn}\right)^2, \left(V_2 - \frac{1}{kn}\right)^2\right]$$

$$= O(n^4)\,\text{Cov}\left[V_1^2 - \frac{2}{kn}V_1, V_2^2 - \frac{2}{kn}V_2\right]$$

$$= O(n^4)\left\{\left[E(V_1^2V_2^2) - \frac{4}{kn}E(V_1^2V_2) + \frac{4}{k^2n^2}E(V_1V_2)\right] - \left[E(V_1^2) - \frac{2}{kn}E(V_1)\right]^2\right\}$$

$$= O(n^4)O(n^{-6}) = O(n^{-2}).$$

If, now, $m_n = n$, it follows from Corollary 1 to Theorem 1 that

$$\frac{1}{\sqrt{n}}\frac{kn(U_n - 1/k)}{\sqrt{(kn-1)(kn+1)^{-1}}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{ni}$$

has a limiting normal distribution with mean zero and variance $1 - 1/k$. The same statement then applies to $k\sqrt{n}(U_n - 1/k)$.

Thus, the sample quantile U_n of order $1/k$ in a sample of $kn - 1$ from a rectangular distribution is asymptotically normal with expected value $1/k$ and variance $(k-1)/k^3n$.

Clearly, an analogous statement holds with $1/k$ replaced by any real number q in $(0, 1)$. This conclusion extends to other distributions than the rectangular, e.g., if the c.d.f. $F(x)$ has a continuous non-zero derivative at the unique solution x_k of $F(x) = 1/k$. These facts are, of course, well known.

Note, in addition, that

$$EX_{n1}X_{n2} = O(n^2)E[(V_1 - 1/kn)(V_2 - 1/kn)]$$

$$= O(n^2)\left[\frac{1}{kn(kn+1)} - \frac{2}{kn}\frac{1}{kn} + \frac{1}{(kn)^2}\right]$$

$$= O(n^2)O(n^{-3}) = o(1).$$

Thus, if (for specificity) $k = 2$, an application of Corollary 3 yields the conclusion that

$$\frac{1}{\sqrt{n}}\left[\frac{2n(U_n - \frac{1}{2})}{\sqrt{(2n-1)(2n+1)^{-1}}} - n\tilde{X}_n\right] = \sqrt{n}\left[2\sqrt{\frac{2n+1}{2n-1}}(\tilde{X}_n - \tfrac{1}{2}) - \bar{X}_n\right]$$

is normally distributed in the limit with mean zero and variance $\frac{1}{2}$ where $\tilde{X}_n$ denotes the sample median. This appears to be new but hardly of overwhelming interest. A comparable result may be demonstrated in the case of a random casting of r_n objects into n cells referred to in Section 3.

EXAMPLE 2. *Ranks.* Let $R_1, \cdots, R_{k_n}$ be a random permutation of the intege (1, 2, $\cdots$, k_n). Define

$$X_{ni} = \frac{R_i - \frac{k_n + 1}{2}}{\sqrt{\frac{k_n^2 - 1}{12}}}.$$

Then, $(R_1, \cdots, R_{k_n})$ and $(X_{n1}, \cdots, X_{n,k_n})$ each comprise a set of i.r.v.'s. Mor over,

$$\sum_{i=1}^{k_n} X_{ni} = 0, \qquad \sum_{i=1}^{k_n} X_{ni}^2 = 1,$$

and

$$\max_{1 \leq i \leq k_n} \frac{|X_{ni}|}{\sqrt{k_n}} \frac{\sqrt{12}}{\sqrt{k_n}} \frac{\left|k_n - \frac{k_n + 1}{2}\right|}{\sqrt{k_n^2 - 1}} = \frac{1}{\sqrt{k_n}} \sqrt{\frac{3(k_n - 1)}{(k_n + 1)}} \to 0.$$

A direct application of Corollary 2 of Theorem 1 yields the limiting normalit (mean 0, variance $1 - \alpha$) of

$$\sum_{i=1}^{m_n} \frac{R_i - \frac{k_n + 1}{2}}{\sqrt{\frac{k_n^2 - 1}{12}}},$$

where $\lim_{n \to \infty} m_n/k_n = \alpha$, $0 < \alpha < 1$, a familiar result.

REFERENCES

[1] E. SPARRE ANDERSEN, "On sums of symmetrically dependent random variables," *Skand. Aktuarietidskr.*, Vol. 36 (1953), pp. 123–138.

[2] E. SPARRE ANDERSEN, "On fluctuations of sums of random variables," I, II, *Math. Skand.*, Vol. 1 (1953), pp. 263–285, and Vol. 2 (1954) pp. 195–223.

[3] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[4] BRUNO DE FINETTI, "La previsions; ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré*, Vol. 7 (1937), pp. 1–68.

[5] GNEDENKO-KOLMOGOROFF, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley Publ. Co., Inc., Cambridge, Mass., 1954.

[6] MICHEL LOÈVE, *Probability Theory*, D. Van Nostrand, New York, 1955.

[7] R. VON MISES, "Uber Aufteilungs- und Besetzungswahrscheinlichkeiten," *Revue de la Faculte des Sciences de l'Universite d'Istanbul N.S.*, (1939), pp. 145–163.

[8] GOTTFRIED NOETHER, On a theorem by Wald and Wolfowitz, *Ann. Math. Stat.*, Vol. 20 (1949), pp. 455–458.

[9] IRVING WEISS, "Limiting Distributions in Some Occupancy Problems," Technical Report No. 28, Applied Mathematics and Statistics Laboratory, Stanford University, Office of Naval Research Contract N6onr-25140, 1955.

LINEAR ESTIMATION FROM CENSORED DATA

By R. L. Plackett

University of Liverpool

1. Introduction. Suppose that a sample of n random variables is taken from a continuous probability distribution, whose density function is $f[(y - \mu)/\sigma]/\sigma$, where μ and σ are unknown. Arrange the variables in order of magnitude, and denote them by $y_1, y_2, \cdots, y_n$, where

$$y_1 < y_2 < \cdots < y_n$$

We shall discuss the problem of estimating μ and σ from the k successive variables $y_u, y_{u+1}, \cdots, y_v$, where $v = u + k - 1$. This problem arises, for example, in life-testing, and some applications are described by Gupta [7]

When using the principal results derived here, the expected values of ordered variables are essential, but tables of these quantities for normal samples are, at present somewhat limited. However, recent studies by Berkson [1] have shown the importance of the logistic distribution, which closely resembles the normal, and some properties of ordered logistic variables are given in Section 2. We now turn to the main problem. If u and v are fixed, the best linear unbiased estimates of μ and σ can be calculated by least squares, given the expected value and dispersion matrix of the vector of ordered variables (Godwin [6], Lloyd [11], Gupta [7]. In general, special tables become necessary, and it seems desirable to obtain simple formulae when samples are moderate or large in size. This is achieved in Section 3, where asymptotic values of the coefficients of $y_u, y_{u+1}, \cdots, y_v$ are derived. An examination of the conditions involved is supplied in Section 4, by considering the limiting form of the maximum likelihood equations. Several illustrative numerical tables complete the paper.

2. Ordered logistic variables. The logistic distribution is defined by

$$L = \log\{p/(1 - p)\}, \tag{1}$$

where p is the probability of a value less than L. Suppose that $L(i; n)$ is the ith variable in ascending order in a sample of size n from this distribution. Then

$$\begin{aligned} \mathcal{E} \exp\{wL(i; n)\} &= \frac{n!}{(i-1)!(n-i)!}\int_0^1 \left(\frac{p}{1-p}\right)^w p^{i-1}(1-p)^{n-i}\,dp \\ &= \frac{(i-1+w)!(n-i-w)!}{(i-1)!(n-i)!}. \end{aligned} \tag{2}$$

Take logarithms, differentiate with respect to w, and put $w = 0$. The cumulants of $L(i; n)$ are

$$\kappa_j(i; n) = \frac{d^j}{dw^j}\log(i-1)! + (-1)^j\frac{d^j}{dw^j}\log(n-i)! \tag{3}$$

Received July 1, 1957; revised September 17, 1957.

and are thus expressible in terms of polygamma functions, tabulated for $j = 1, 2, 3, 4$ in [2]. For $(i - 1) > (n - i)$, we obtain

$$\kappa_1(i; n) = \frac{1}{(n - i + 1)} + \frac{1}{(n - i + 2)} + \cdots + \frac{1}{(i - 1)}, \tag{4}$$

$$\kappa_2(i; n) = \frac{\pi^2}{3} - \left\{1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{(i - 1)^2}\right\} - \left\{1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{(n - i)^2}\right\}, \tag{5}$$

$$\kappa_3(i; n) = 2\left\{\frac{1}{(n - i + 1)^3} + \frac{1}{(n - i + 2)^3} + \cdots + \frac{1}{(i - 1)^3}\right\}, \tag{6}$$

$$\kappa_4(i; n) = \frac{2\pi^4}{15} - 6\left\{1 + \frac{1}{2^4} + \frac{1}{3^4} + \cdots + \frac{1}{(i - 1)^4}\right\} - 6\left\{1 + \frac{1}{2^4} + \frac{1}{3^4} + \cdots + \frac{1}{(n - i)^4}\right\}. \tag{7}$$

Suppose that $x(i; n)$ is the ith variable in ascending order in a sample of size n from the probability distribution whose density function is $f(x)$ and distribution function $F(x)$. Let α be fixed, $0 < \alpha < 1$, and define t by

$$\alpha = F(t). \tag{8}$$

We require the two following results. As $n \to \infty$, with $i = [n\alpha]$

$$\mathcal{E}x(i; n) = t + O(n^{-1}) \tag{9}$$

and

$$F\{\mathcal{E}x(i + 1; n)\} - F\{\mathcal{E}x(i; n)\} = 1/n + O(n^{-2}). \tag{10}$$

The proofs are based on the Taylor expansion of x, considered as a function of L, about the value $L = \kappa_1(i; n)$. This, after expectation, gives

$$\mathcal{E}x(i; n) = x^{(0)} + \tfrac{1}{2}x^{(2)}\kappa_2 + \tfrac{1}{6}x^{(3)}\kappa_3 + \tfrac{1}{24}x^{(4)}(\kappa_4 + 3\kappa_2^2) + \cdots, \tag{11}$$

where $x^{(j)}$ is the value at $L = \kappa_1(i; n)$ of the jth derivative of x with respect to L. Now

$$\kappa_1(i; n) = \tfrac{1}{2}\log\{(i - 1)i/(n - i)(n - i + 1)\} + O(n^{-2}) \tag{12}$$

whence

$$\kappa_1(i; n) = \lambda + O(n^{-1}), \tag{13}$$

where

$$\lambda = \log\{\alpha/(1 - \alpha)\}. \tag{14}$$

Also

$$\kappa_j(i; n) = O(n^{1-j}) \qquad (j = 2, 3, \cdots). \tag{15}$$

Assuming $x^{(2)}$ to be bounded, we can substitute (13) and (15) in (11) to obtain (9). As regards (10), we suppose that $x^{(2)}$ and $x^{(3)}$ are bounded, in which case

$$(16) \qquad \mathcal{E}x(i+1;n) - \mathcal{E}x(i;n) = \left\{\frac{1}{i} + \frac{1}{(n-i)}\right\} x^{(1)} + O(n^{-2}).$$

On the further assumption that df/dx is bounded, (10) results.

We shall now consider the standard normal distribution in more detail. Denote its density function by $\phi(x)$ and distribution function by $\Phi(x)$. Here

$$(17) \quad x^{(1)} = \Phi(1-\Phi)/\phi,$$

$$(18) \quad x^{(2)} = x^{(1)}\{xx^{(1)} - (2\Phi - 1)\},$$

$$(19) \quad x^{(3)} = (x^{(1)})^3 + 2xx^{(1)}x^{(2)} + x^{(2)}(1-2\Phi) - 2x^{(1)}\Phi(1-\Phi),$$

$$(20) \quad x^{(4)} = 5(x^{(1)})^2x^{(2)} + x^{(3)}\{2xx^{(1)} - (2\Phi-1)\} + 2x^{(2)}\{xx^{(2)} - 2\Phi(1-\Phi)\} + 2x^{(1)}(2\Phi-1)\Phi(1-\Phi).$$

These derivatives are all bounded, their maximum absolute values being given below.

$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$
0 62666	0.07376	0 06724	0 04597

The absolute value of the remainder after $(j-1)$ terms of the series (11) is at most $\beta_j \max|x^{(j)}|/j!$, where β_j is the jth absolute moment about the mean of the ith ordered logistic variable in a sample of n. Since β_j is known when j is even and the inequality $(\beta_j)^{1/j} \leqq (\beta_{j+1})^{1/(j+1)}$ is available when j is odd, we can thus assign bounds to $\mathcal{E}x(i;n)$ for all values of j. As an illustration, take $\mathcal{E}x(19;25)$.

j	Series (11) to j terms	Absolute maximum error
1	0.642835	0.007656
2	0 636781	0.002521
3	0.636656	0.000262

David and Johnson [5] express x as a function of Φ, and the value for $\mathcal{E}x(19;25)$ from the first four terms of the series on p. 236 of their paper is 0.636904. However, their formula is arranged as a power series in $(n+2)^{-1}$, and a similar rearrangement of (11) would be necessary before a full comparison of the two approaches can be made. This will be undertaken on another occasion.

3. Least squares estimation. Let t_i denote the expectation of $(y_i - \mu)/\sigma$. Write

$$(21) \qquad f_i = f(t_i),$$

$$(22) \qquad p_i = F(t_i),$$

and

(23) $$q_i = 1 - p_i \qquad (i = u, u + 1, \cdots, v).$$

Let m be the vector of $(y_i - \mu)/\sigma$ for $i = u, u + 1, \cdots, v$ and put

(24) $$t = \mathcal{E}m$$

and

(25) $$V = \mathfrak{D}m = \mathcal{E}\{(m - \mathcal{E}m)(m - \mathcal{E}m)'\}.$$

In principle, t and V can be computed from the known function $f(x)$. The estimate of

(26) $$\theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}$$

given by generalized least squares is

(27) $$\theta^* = (A'V^{-1}A)^{-1}A'V^{-1}y,$$

where

(28) $$A = [1 \quad t].$$

As V is difficult to handle analytically, we replace it by W, a symmetric matrix whose elements are $\{a_ib_j\}$ for $i \leqq j$, where

(29) $$a_i = p_i/f_i \qquad (i = u, u + 1, \cdots, v)$$

and

(30) $$b_j = q_j/f_j \qquad (j = u, u + 1, \cdots, v).$$

Since $\mathfrak{D}y \sim W\sigma^2/n$, the unbiased estimate

(31) $$\theta^+ = (A'W^{-1}A)^{-1}A'W^{-1}y$$

may be presumed to have the same asymptotic properties as θ^*. We therefore consider the limiting form of θ^+.

The inverse of W has been given by Hammersley and Morton [9]. Put

(32) $$a_{u-1} = 0, \qquad a_{v+1} = 1, \qquad b_{u-1} = 1, \qquad b_{v+1} = 0,$$

Then

(33) $$W^{-1} = \begin{bmatrix} c_u & d_u & 0 & 0 & \cdots & 0 \\ d_u & c_{u+1} & d_{u+1} & 0 & \cdots & 0 \\ 0 & d_{u+1} & c_{u+2} & d_{u+2} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & & c_{v-1} & d_{v-1} \\ 0 & 0 & \cdots & & d_{v-1} & c_v \end{bmatrix},$$

where

(34) $$c_i = (a_{i+1}b_{i-1} - a_{i-1}b_{i+1}) \,/\, (a_ib_{i+1} - a_{i+1}b_i)(a_{i-1}b_i - a_ib_{i-1})$$

and

(35) $$d_i = 1/(a_i b_{i+1} - a_{i+1} b_i).$$

Denote $A'W^{-1}$ by G, and define

(36) $$h_s = \tfrac{1}{2}(p_{s+1} - p_{s-1}) \qquad (s = u + 1, u + 2, \cdots, v - 1),$$

(37) $$h_u = p_{u+1} - p_u ,$$

and

(38) $$h_v = p_v - p_{v-1} .$$

If the elements of G are considered as functions of $p_u, p_{u+1}, \cdots, p_v$ then g_{1u} depends on p_u and p_{u+1}; g_{1s} on p_{s-1}, p_s, and p_{s+1}; and g_{1v} on p_{v-1} and p_v. In g_{1u}, put $p_{u+1} = p_u + h_u$; in g_{1s}, replace p_{s-1} by $p_s - h_s$ and p_{s+1} by $p_s + h_s$; and in g_{1v}, put $p_{v-1} = p_v - h_v$. The first and third substitutions are exact; the second one is approximate, but if $n \to \infty$ with $u = [n\alpha]$ and $v = [n\beta]$, the values of p_i tend to become equally spaced between α and β, as (10) shows. The elements in the second row of G are treated similarly, so that both elements in the ith column are now expressed as functions of p_i and h_i. Expanding by Taylor series as far as h_s^3 in the numerators of g_{1s} and g_{2s} and as far as h^2 elsewhere, the elements of G finally reduce, after a good deal of straightforward algebra, to the expressions given below. The primes signify differentiation with respect to p, so that

(39) $$f' = \frac{d \log f}{dx}$$

and

(40) $$ff'' = \frac{d^2 \log f}{dx^2}.$$

In calculating the elements of $A'W^{-1}A$, we pass from sums involving h to integrals involving dp.

4. Maximum likelihood estimation. The likelihood of $y_u, y_{u+1}, \cdots, y_v$ is

(41) $$\frac{n!}{(u-1)!(n-v)!}\left\{F\left(\frac{y_u - \mu}{\sigma}\right)\right\}^{u-1} \prod_{i=u}^{v} \frac{1}{\sigma} f\left(\frac{y_i - \mu}{\sigma}\right)\left\{1 - F\left(\frac{y_v - \mu}{\sigma}\right)\right\}^{n-v}.$$

Denote by $\hat{\mu}$ and $\hat{\sigma}$ the maximum likelihood estimates of μ and σ, respectively. They satisfy the equations

(42) $$-\frac{(u-1)\hat{\sigma} f\left(\dfrac{y_u - \hat{\mu}}{\hat{\sigma}}\right)}{nF\left(\dfrac{y_u - \hat{\mu}}{\hat{\sigma}}\right)} - \frac{\hat{\sigma}}{n}\sum_{i=u}^{v} \frac{d \log f}{dx}\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}}\right) + \frac{(n-v)\hat{\sigma} f\left(\dfrac{y_v - \hat{\mu}}{\hat{\sigma}}\right)}{n\left\{1 - F\left(\dfrac{y_v - \hat{\mu}}{\hat{\sigma}}\right)\right\}} = 0,$$

TABLE 1

Asymptotic value of $A' W^{-1}$

Row—Column	General Density	Normal Density
(1, u)	$f_u^2/p_u - f'_u f_u - \frac{1}{2}h_u f_u f''_u$	$f_u^2/p_u + t_u f_u + \frac{1}{2}h_u$
(1, s)	$-h_s f_s f''_s$	h_s
(1, v)	$f_v^2/q_v + f'_v f_v - \frac{1}{2}h_v f_v f''_v$	$f_v^2/q_v - t_v f_v + \frac{1}{2}h_v$
(2, u)	$t_u f_u^2/p_u - t_u f'_u f_u - f_u - \frac{1}{2}h_u(f'_u + f_u f''_u t_u)$	$t_u f_u^2/p_u + t_u^2 f_u - f_u + h_u t_u$
(2, s)	$-h_s(f'_s + f_s f''_s t_s)$	$2h_s t_s$
(2, v)	$t_v f_v^2/q_v + t_v f'_v f_v + f_v - \frac{1}{2}h_v(f'_v + f_v f''_v t_v)$	$t_v f_v^2/q_v - t_v^2 f_v + f_v + h_v t_v$

TABLE 2

Asymptotic value of $A' W^{-1} A$

Row—Column	General Density	Normal Density
(1, 1)	$-\int_{p_u}^{p_v} ff''\, dp + f_v^2/q_v + f_v f'_v + f_u^2/p_u - f_u f'_u$	$f_v^2/q_v - t_v f_v + p_v + f_u^2/p_u + t_u f_u - p_u$
(1, 2) and (2, 1)	$-\int_{p_u}^{p_v} tff''\, dp + t_v f_v^2/q_v + t_v f_v f'_v + t_u f_u^2/p_u - t_u f_u f'_u$	$t_v f_v^2/q_v - t_v^2 f_v - f_v + t_u f_u^2/p_u + t_u^2 f_u + f_u$
(2,2)	$-\int_{p_u}^{p_v} t^2 ff''\, dp + t_v^2 f_v^2/q_v + t_v^2 f_v f'_v + p_v + t_u^2 f_u^2/p_u - t_u^2 f_u f'_u - p_u$	$t_v^2 f_v^2/q_v - t_v^3 f_v - t_v f_v + 2p_v + t_u^2 f_u^2/p_u + t_u^3 f_u + t_u f_u - 2p_u$

(43)
$$-\frac{(u-1)(y_u-\hat{\mu})f\left(\dfrac{y_u-\hat{\mu}}{\hat{\sigma}}\right)}{nF\left(\dfrac{y_u-\hat{\mu}}{\hat{\sigma}}\right)} - \frac{k\hat{\sigma}}{n} - \frac{1}{n}\sum_{i=u}^{v}(y_i-\hat{\mu})\frac{d\log f}{dx}\left(\frac{y_i-\hat{\mu}}{\hat{\sigma}}\right) + \frac{(n-v)(y_v-\hat{\mu})f\left(\dfrac{y_v-\hat{\mu}}{\hat{\sigma}}\right)}{n\left\{1-F\left(\dfrac{y_v-\hat{\mu}}{\hat{\sigma}}\right)\right\}} = 0,$$

where

$$\frac{d\log f}{dx}\left(\frac{y_i-\hat{\mu}}{\hat{\sigma}}\right)$$

means the value at $(y_i - \hat{\mu})/\hat{\sigma}$ of the function $d\log f/dx$. The direct solution of (42) and (43) for normal samples has been described by Cohen [3], who used successive approximation; and, when $u = 1$, by Gupta [7], who calculated a special table which shortens the work. Halperin [8] has indicated conditions under which

(a) the maximum likelihood equations have a consistent set of solutions $\hat{\mu}$, $\hat{\sigma}$;

(b) $\sqrt{n}(\hat{\mu} - \mu)$ and $\sqrt{n}(\hat{\sigma} - \sigma)$ have a bivariate normal limit distribution;
(c) the dispersion matrix of the limit distribution is best in the sense of Cramér [4], §32.6.

The necessary assumptions involve derivatives of $f[(y - \mu)/\sigma]/\sigma$ with respect to μ and σ, and we shall suppose henceforth that they are satisfied.

We expand $(y_i - \hat{\mu})/\hat{\sigma}$ in a Taylor series about t_i and obtain

$$(44)\quad \begin{aligned} &-\frac{(u-1)\hat{\sigma}}{n}\left\{\frac{f_u}{p_u} + \left(\frac{y_u - \hat{\mu}}{\hat{\sigma}} - t_u\right)\left(\frac{f_u f_u'}{p_u} - \frac{f_u^2}{p_u^2}\right) + \frac{1}{2}\left(\frac{y_u - \hat{\mu}}{\hat{\sigma}} - t_u\right)^2 C\right\} \\ &-\frac{\hat{\sigma}}{n}\sum_{i=u}^{v}\left\{f_i' + \left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)f_i f_i'' + \frac{1}{2}\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)^2 D_i\right\} \\ &+\frac{(n-v)\hat{\sigma}}{n}\left\{\frac{f_v}{q_v} + \left(\frac{y_v - \hat{\mu}}{\hat{\sigma}} - t_v\right)\left(\frac{f_v f_v'}{q_v} + \frac{f_v^2}{q_v^2}\right) + \frac{1}{2}\left(\frac{y_v - \hat{\mu}}{\hat{\sigma}} - t_v\right)^2 E\right\} = 0, \end{aligned}$$

$$(45)\quad \begin{aligned} &-\frac{(u-1)\hat{\sigma}}{n}\left\{\frac{t_u f_u}{p_u} + \left(\frac{y_u - \hat{\mu}}{\hat{\sigma}} - t_u\right)\left(\frac{f_u}{p_u} + \frac{t_u f_u f_u'}{p_u} - \frac{t_u f_u^2}{p_u^2}\right)\right. \\ &\left.+\frac{1}{2}\left(\frac{y_u - \hat{\mu}}{\hat{\sigma}} - t_u\right)^2 R\right\} - \frac{k\hat{\sigma}}{n} - \frac{\hat{\sigma}}{n}\sum_{i=u}^{v}\left\{t_i f_i'\right. \\ &\left.+\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)(t_i f_i f_i'' + f_i') + \frac{1}{2}\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)^2 S_i\right\} + \frac{(n-v)\hat{\sigma}}{n} \\ &\left\{\frac{t_v f_v}{q_v} + \left(\frac{y_v - \hat{\mu}}{\hat{\sigma}} - t_v\right)\left(\frac{f_v}{q_v} + \frac{t_v f_v f_v'}{q_v} + \frac{t_v f_v^2}{q_v^2}\right) + \frac{1}{2}\left(\frac{y_v - \hat{\mu}}{\hat{\sigma}} - t_v\right)^2 T\right\} = 0. \end{aligned}$$

Here C, D_i, E, R, S_i and T are second-order derivatives with respect to x evaluated at points intermediate between $(y_i - \hat{\mu})\hat{\sigma}$ and t_i; and the primes have their previous meaning.

We assume that the second-order derivatives of

$$\frac{d \log f}{dx}, \quad x\frac{d \log f}{dx}, \quad \frac{f}{p}, \quad \frac{f}{q}, \quad \frac{xf}{p}, \quad \frac{xf}{q},$$

with respect to x, are functions of bounded variation. This condition is not satisfied if $F(x) = 0$ at a finite value of x, since then

$$\left|\frac{d^2}{dx^2}\left(\frac{f}{p}\right)\right| \to \infty$$

at the lower terminus of the distribution; nor if $F(x) = 1$ for finite x, since

$$\left|\frac{d^2}{dx^2}\left(\frac{f}{q}\right)\right| \to \infty$$

there. However, all is well for the normal and logistic distributions, as the following table shows.

Maximum absolute values of second-order derivatives

Distribution	$\frac{d \log f}{dx}$	$x \frac{d \log f}{dx}$	$\frac{f}{p}$	$\frac{f}{q}$	$\frac{xf}{p}$	$\frac{xf}{q}$
Normal	0	2.00	0.30	0.30	2.00	2.00
Logistic	0.19	1.00	0.10	0.10	0.50	0.50

Let α and β be fixed such that $0 < \alpha < \beta < 1$. We assume that $f(t) \geqq c > 0$ wherever $F^{-1}(\alpha) \leqq t \leqq F^{-1}(\beta)$. For any such t, let f, p and q be defined in accordance with (21), (22), and (23). Then

(46) $$z^2 = pq/f^2$$

has a finite maximum, which we denote by z_1^2. We proceed to derive the form taken by the maximum likelihood equations when $u = [n\alpha]$, $v = [n\beta]$, and $n \rightarrow \infty$.

Consider the variable

(47) $$\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right) = \frac{(y_i - \mu - t_i \sigma) - (\hat{\mu} - \mu) - t_i(\hat{\sigma} - \sigma)}{\hat{\sigma}}$$

Given $\epsilon_1 > 0$, $\epsilon_2 > 0$, and ϵ_3 such that $\sigma > \epsilon_3 > 0$,

(48) $$\Pr\left\{\left|\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right| > \frac{\epsilon_1 + \epsilon_2}{\hat{\sigma} - \epsilon_3}\right\} < \Pr\{|y_i - \mu - t_i\sigma| > \epsilon_1\} + \Pr\{|(\hat{\mu} - \mu) + t_i(\hat{\sigma} - \sigma)| > \epsilon_2\} + \Pr\{|\hat{\sigma} - \sigma| > \epsilon_3\}.$$

Typically, $i = [np]$, and as $n \rightarrow \infty$, $\frac{\sqrt{n}}{z}\left(\frac{y_i - \mu}{\sigma} - t\right)$ is asymptotically normal with zero mean and unit variance (Cramér [4], §28.5). According to (9), $(t_i - t)$ is $O(n^{-1})$ and so $\frac{\sqrt{n}}{z}\left(\frac{y_i - \mu}{\sigma} - t_i\right)$ has the same limit distribution. Hence

(49) $$\Pr\{|y_i - \mu - t_i\sigma| > \epsilon_1\} \sim 2\Phi(-n^{1/2}\epsilon_1/\sigma z) \sim 2(\sigma z/n^{1/2}\epsilon_1)\phi(n^{1/2}\epsilon_1/\sigma z).$$

Similarly, by the asymptotic properties of $\hat{\mu}$ and $\hat{\sigma}$, there exist finite quantities z_2 and z_3 such that

(50) $$\Pr\{|(\hat{\mu} - \mu) + t_i(\hat{\sigma} - \sigma)| > \epsilon_2\} \sim 2(\sigma z_2/n^{1/2}\epsilon_2)\phi(n^{1/2}\epsilon_2/\sigma z_2)$$

and

(51) $$\Pr\{|\hat{\sigma} - \sigma| > \epsilon_3\} \sim 2(\sigma z_3/n^{1/2}\epsilon_3)\phi(n^{1/2}\epsilon_3/\sigma z_3).$$

Consequently, as $n \rightarrow \infty$,

(52) $$\sum_{i=u}^{v} \Pr\left\{\left|\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right| > \frac{\epsilon_1 + \epsilon_2}{\sigma - \epsilon_3}\right\} < 2(\beta - \alpha)n^{1/2}\sigma \sum_{j=1}^{3} (z_j/\epsilon_j)\phi(n^{1/2}\epsilon_j/\sigma z_j).$$

Thus, given $\epsilon > 0$ and $\delta > 0$, arbitrarily small, we can find N such that

$$(53) \qquad \sum_{i=u}^{v} \Pr\left\{\left|\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right| > \epsilon\right\} < \delta \qquad \text{for} \quad n \geqq N.$$

Therefore, by Boole's inequality,

$$(54) \qquad \Pr\left\{\left|\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right| \leqq \epsilon \quad \text{for} \quad i = u, u+1, \cdots, v\right\} \geqq 1 - \delta \quad \text{for} \quad n \geqq N.$$

When this event occurs, the terms in (44) and (45) which involve $\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} - t_i\right)^2$ are negligible compared with the remaining terms, because C, D_i, E, R, S_i and T are all bounded. We therefore omit the squared terms, and replace sums by integrals as before (cf. Hoeffding [10]), at the same time making an Euler-Maclaurin adjustment on the coefficients of y_u and y_v, so as to correct for bias in the estimates and bring the results into line with those previously obtained. The linearized form taken asymptotically by the maximum likelihood equations is then given by the coefficients in Tables 1 and 2 on replacing h_i by $1/n$ throughout. We shall use μ^0 and σ^0 to denote the corresponding linearized estimates. The asymptotic dispersion matrix of the maximum likelihood estimates is σ^2/n times the inverse of the equations for μ^0 and σ^0.

These results show, not only that the maximum likelihood estimates of μ and σ are asymptotically linear, but also that the best linear unbiased estimates

TABLE 3A

Coefficients of ordered variables when estimating the mean. θ^* above, θ° below

k	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y	y_9	y_{10}
2	−1.8634	2.8634								
	−2.1547	3.0554								
3	−0.6596	−0.2138	1.8734							
	−0.7487	−0.2248	1.9309							
4	−0.2923	−0.0709	0.0305	1.3327						
	−0.3304	−0.0780	0.0362	1.3543						
5	−0.1240	−0.0016	0.0549	0.0990	0.9718					
	−0.1418	−0.0055	0.0567	0.1071	0.9797					
6	−0.0316	0.0383	0.0707	0.0962	0.1185	0.7078				
	−0.0394	0.0366	0.0718	0.1003	0.1261	0.7100				
7	0.0244	0.0636	0.0818	0.0962	0.1089	0.1207	0.5045			
	0 0222	0.0633	0.0829	0 0988	0.1131	0.1270	0.5045			
8	0.0605	0.0804	0.0898	0.0972	0.1037	0.1099	0.1161	0.3424		
	0.0616	0.0812	0.0911	0.0992	0.1066	0.1137	0.1210	0.3423		
9	0.0843	0.0921	0.0957	0.0986	0.1011	0.1036	0.1060	0.1085	0.2101	
	0.0877	0.0934	0.0973	0.1004	0.1033	0.1060	0.1089	0.1120	0.2113	
10	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000
	0.1048	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1018	0.1048

are asymptotically normal and efficient. In order to compute the coefficients of the ordered variables, only tables of $f(x)$, $F(x)$ and t_i are necessary. For normal samples, Teichroew [14] gives t_i to 10 D for $n \leqq 20$, and an extension to $n \leqq 100$ with 24 D is being prepared (Ruben [12]). For logistic samples, explicit formulae have already been given.

5. Numerical tables. Tables 3A and 3B refer to the estimation of θ from the smallest k observations in a sample of size $n = 10$ from a normal distribution. They give the coefficients of $y_1, y_2, \cdots, y_k$ for $k \leqq 10$ in

(i) the best linear unbiased estimate, $\theta^* = \begin{bmatrix} \mu^* \\ \sigma^* \end{bmatrix}$,

(ii) the linearized maximum likelihood estimate, $\theta^0 = \begin{bmatrix} \mu^\circ \\ \sigma^\circ \end{bmatrix}$

Table 4 gives the coefficients of μ and σ in the expectation of θ^0. Suppose in general that

$$\mathcal{E}\theta^0 = B\theta \tag{55}$$

Then $B^{-1}\theta^0$ is an unbiased estimate of θ, and the efficiencies of its elements, relative to μ^* and σ^* respectively, have been calculated from the table of $\mathfrak{D}m$ in Sarhan and Greenberg [13] when, as above, $n = 10$, $u = 1$, and $v = 2, 3, \cdots 10$. These efficiencies never fall below 0.9998, a result which suggests that θ^0, corrected for bias, can be used in place of θ^*, with negligible loss of efficiency, for all sample sizes of practical importance.

TABLE 3B

Coefficients of ordered variables when estimating the standard deviation. θ^* above, θ° below

k	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
2	−1.8608	1.8608								
	−2.1366	2.0404								
3	−0.9625	−0.4357	1.3981							
	−1.0767	−0.4586	1.4738							
4	−0.6520	−0.3150	−0.1593	1.1263						
	−0.7190	−0.3330	−0.1611	1.1681						
5	−0.4419	−0.2491	−0.1362	−0.0472	0.9243					
	−0.5374	−0.2631	−0.1414	−0.0425	0.9499					
6	−0.3931	−0.2063	−0.1192	−0.0501	0.0111	0.7576				
	−0.4266	−0.2175	−0.1250	−0.0498	0.0180	0.7740				
7	−0.3252	−0.1758	−0.1058	−0.0502	−0.0006	0.0469	0.6107			
	−0.3513	−0.1849	−0.1114	−0.0517	0.0022	0.0545	0.6218			
8	−0.2753	−0.1523	−0.0947	−0.0488	−0.0077	0.0319	0.0722	0.4746		
	−0.2963	−0.1600	−0.0998	−0.0510	−0.0069	0.0358	0.0799	0.4830		
9	−0.2364	−0.1334	−0.0851	−0.0465	−0.0119	0.0215	0.0559	0.0936	0.3423	
	−0.2539	−0.1399	−0.0897	−0.0490	−0.0122	0.0234	0.0602	0.1009	0.3505	
10	−0.2044	−0.1172	−0.0763	−0.0436	−0.0142	0.0142	0.0436	0.0763	0.1172	0.2044
	−0.2196	−0.1231	−0.0807	−0.0462	−0.0151	0.0151	0.0462	0.0807	0.1231	0.2196

TABLE 4

Expectation of θ^0

k	μ^0		σ^0	
	μ	σ	μ	σ
2	0.9007	0.2560	−0.0962	1.2446
3	0.9574	0.1104	−0.0616	1.1492
4	0.9821	0.0539	−0.0450	1.1066
5	0.9962	0.0260	−0.0346	1.0827
6	1.0054	0.0108	−0.0270	1.0678
7	1.0119	0.0022	−0.0208	1.0583
8	1.0166	−0.0023	−0.0153	1.0529
9	1.0204	−0.0038	−0.0097	1.0523
10	1.0243	0.0000	0.0000	1.0668

TABLE 5

$1/n$	h_1	h_2	h_3	h_4	h_5
0.5000	0.4274				
0.3333	0.3013	0.3013			
0.2500	0.2316	0.2326			
0.2000	0.1879	0.1888	0.1897		
0.1667	0.1580	0.1588	0.1597		
0.1429	0.1362	0.1370	0.1378	0.1378	
0.1250	0.1198	0.1204	0.1212	0.1212	
0.1111	0.1068	0.1074	0.1081	0.1082	0.1082
0.1000	0.0964	0.0970	0.0976	0.0976	0.0976

Table 5 also refers to normal samples. Used in conjunction with the relation

(56) $$h_i = h_{n+1-i},$$

it gives the values of h_i for $n = 2, 3, \cdots 10$ and $1 \leqq i \leqq n$. That there is close agreement between θ^+ and θ^0 can be inferred from Table 5 in particular and (10) in general.

6. Acknowledgements. I am grateful to Mr. C. J. Taylor for doing all the calculations; and to the referee for a correction to my maximal remainder in (11), and other helpful remarks.

REFERENCES

[1] J. Berkson, "Why I prefer logits to probits," *Biometrics*, Vol. 7 (1951), pp. 327–339.
[2] *British Association Mathematical Tables*, Vol. 1 (1931).
[3] A. C. Cohen, Jr., "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 557–569.
[4] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[5] F. N. David and N. L. Johnson, "Statistical treatment of censored data. Part 1. Fundamental formulae," *Biometrika*, Vol. 41 (1954), pp. 228–240.

[6] H. J. Godwin, "On the estimation of dispersion by linear systematic statistics," *Biometrika*, Vol. 36 (1949), pp. 92–100.
[7] A. K. Gupta, "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, Vol. 39 (1952), pp. 260–273.
[8] M. Halperin, "Maximum likelihood estimation in truncated samples," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 226–238.
[9] J. M. Hammersley and K. W. Morton, "The estimation of location and scale parameters from grouped data," *Biometrika*, Vol. 41 (1954), pp. 296–301.
[10] W. Hoeffding, "On the distribution of the expected values of the order statistics," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 93–100.
[11] E. H. Lloyd, "Least squares estimation of location and scale parameters using order statistics," *Biometrika*, Vol. 39 (1952), pp. 88–95.
[12] H. Ruben, Personal communication.
[13] A. E. Sarhan and B. G. Greenberg, "Estimation of location and scale parameters from singly and doubly truncated samples. Part I. The normal distribution up to samples of size 10," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 427–451.
[14] D. Teichroew, "Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 410–426.

corresponding to each set T, there exists an integrable function $\lambda(T \mid x)$ on the real line such that

$$P\{f^{-1}(S) \cap g^{-1}(T); \theta\} \underset{\theta}{\equiv} \int_S \lambda(T \mid x)\, dPf^{-1}(x - \theta).$$

This gives us

$$\begin{aligned} P\{f^{-1}(S) \cap g^{-1}(T); 0\} &= P\{f^{-1}(S + \theta) \cap g^{-1}(T + \theta); \theta\} \\ &= \int_S \lambda(T + \theta \mid x + \theta)\, dPf^{-1}(x). \end{aligned}$$

It follows that for every set T, we have

$$\begin{aligned} P\{f^{-1}(S) \cap g^{-1}(T); 0\} &= \int_S \lambda(T - x \mid 0)\, dPf^{-1}(x) \\ &= \int_S \mu(T - x)\, dPf^{-1}(x), \end{aligned}$$

so that

$$\Pr\{g(\omega) \,\varepsilon\, T \mid f(\omega) = x\} = \mu(T - x) \text{ for a.e. } x\ [Pf^{-1}],$$

and consequently

$$\Pr\{g(\omega) - f(\omega) \,\varepsilon\, T \mid f(\omega) = x\} = \mu(T)$$

is independent of x.

COROLLARY 2.1. *Let $t_j(n; x_1, \cdots, x_n), j = 0, 1$, be functions satisfying Assumption* I *and let $t_0(n; x_1, \cdots, x_n)$ be a sufficient statistic for the family of distributions $\prod_{j=1}^{n} F(x_j - \theta), -\infty < \theta < \infty$. Then for any n, if $X_1, \cdots, X_n$ are independent random variables having the common* c.d.f $F[(x - \theta) / \sigma]$, *the random variables $t_0(n; X_1, \cdots, X_n)$ and $t_1(n; X_1, \cdots, X_n) - t_0(n; X_1, \cdots, X_n)$ are independent.*

COROLLARY 2.2. *If $t_0(n; x_1, \cdots, x_n)$ is as in Corollary* 2.1 *and $s(n; x_1, \cdots, x_n)$ is any function satisfying Assumption* II, *the random variables $t_0(n; X_1, \cdots, X_n)$ and $s(n; X_1, \cdots, X_n)$ are independent.*

LEMMA 3[1]. *Let $t_0, t_1, X_1, \cdots, X_n$ be as in Corollary* 2.1 *and suppose that t_0, t_1 also satisfy Assumption* III *with respective sequences $k_0(n), k_1(n)$. Then*

$$k_0(n) \geqq k_1(n), \tag{32}$$

the equality holding if and only if

$$\Pr\{t_0(n; X_1, \cdots, X_n) = t_1(n; X_1, \cdots, X_n)\} = 1. \tag{33}$$

Further, for any a, b such that $a < 0 < b$,

$$\begin{aligned} &\Pr\{a < t_0(n; X_1, \cdots, X_n) - \theta < b\} \\ &\qquad \geqq \Pr\{a < t_1(n; X_1, \cdots, X_n) - \theta < b\}. \end{aligned} \tag{34}$$

[1] It may be of interest to compare this with the results of Pitman [3], pp. 401–402.

Proof. From Assumptions I and III, it follows that

$$\Pr\{t_j(n; X_1, \cdots, X_n) \leqq x\} = F\left\{\frac{(x-\theta)}{\sigma} k_j(n)\right\}, \qquad j = 0, 1. \tag{35}$$

Let $f(z) = \int e^{izx}\, dF(x)$, $\quad c_n = k_0(n)/k_1(n)$, and

$$g\{z/k_0(n)\} = E\{e^{iz[t_1(n;\, Y_1, \cdots, Y_n) - t_0(n;\, Y_1, \cdots, Y_n)]}\}.$$

Then from Corollary 2.1 and (35), we get

$$f(c_n z) = f(z)g(z). \tag{36}$$

First suppose $f(z_0) = 0$ for some z_0. Then $f(c_n^r z_0) = 0$ for every r, and hence $c_n \nless 1$. On the other hand, if $f(z) \neq 0$ for all z, $g(z) = f(c_n z)/f(z)$ is a characteristic function. Hence $f(c_n^r z)/f(z) = g(z)g(c_n z) \cdots g(c_n^{r-1} z)$, $r = 1, 2, \cdots$, is a sequence of characteristic functions. If $c_n < 1$, $\lim_{r\to\infty} f(c_n^r z)/f(z) = 1/f(z)$ for every finite z, and since the limit is continuous in z, it is a characteristic function. But $f(z)$ is also a characteristic function. This implies $|f(z)| = 1$ for all z, which is impossible if F has more than one point of increase. Consequently, $c_n \geqq 1$, that is to say (32) holds.

It follows from (36) that $c_n = 1$ if and only if $g(z) = 1$, in which case (33) follows on account of Assumption I.

Finally, (34) is an immediate consequence of (32) and (35).

Theorem 1. *Let t_1 and s be statistics satisfying Assumptions* I–VI, *and let t_0 be a statistic, satisfying Assumptions* I, III *and* V, *which is sufficient for the family of distributions $\Pi_1^n F(x_j - \theta)$, $-\infty < \theta < \infty$. Let N_i denote the sample-size in the two-stage procedure using (t_i, s), $i = 0, 1$, with the same m, α, β, and δ. Then*

$$E(N_0) \leqq E(N_1) \quad \textit{for} \quad \textit{all} \quad \sigma, \tag{37}$$

the equality holding for all σ if and only if

$$\Pr\{t_0(n; X_1, \cdots, X_n) = t_1(n; X_1, \cdots, X_n)\} = 1 \textit{ for all } n \geqq m.$$

Proof. The hypotheses of the theorem and Corollary 2.2 enable us to use (t_0, s) for the procedure described in the previous section. With the notation used there we have from (30), $G_i(x) = F(x)$, $i = 0, 1$, and from (9),

$$\rho_0 = \rho_1 . \tag{38}$$

From (13), we get

$$E(N_i) = m + \sum_{r=m}^{\infty} \bar{H}\{k_i(r)\sigma^{-1}\rho_i^{-1}; m\},$$

and using (32) and (38) the result follows.

Remark. Theorem 1 solves part of the problem of optimization of the two-sample procedure by showing that if a suitable sufficient estimator of θ exists, it is the best "t" to use. This leaves us with the problem of choosing "s". We shall see that in the case of the normal and exponential distributions the best pair (t, s) to use, asymptotically as $\sigma \to \infty$, is the pair of sufficient statistics (t_0, s_0).

LEMMA 4[2] *Let* $s_0(n; x_1, \cdots, x_n)$ *and* $s(n; x_1, \cdots, x_n)$ *be statistics satisfying Assumptions* II *and* VI, *and let* t_0 *be a statistic satisfying Assumption* I. *Let* (t_0, s_0) *be sufficient for the family* $\Pi_1^n F[(x_i - \theta)/\sigma]$, $-\infty < \theta < \infty$, $\sigma > 0$. *Then* $t_0(n; X_1, \cdots, X_n)$, $s_0(n; X_1, \cdots, X_n)$, *and* $s(n; X_1, \cdots, X_n)/s_0(n; X_1, \cdots, X_n)$ *are mutually independent.*

PROOF. This result can be proved formally along the lines used for Lemma 2, but this seems hardly necessary, and only an outline in terms of conditional probabilities will be given.

Let $u(n; x_1, \cdots, x_n) = s(n; x_1, \cdots, x_n)/s_0(n; x_1, \cdots, x_n)$, and note that u is invariant under the transformation $x_i \to \sigma x_i + \theta$, $i = 1, \cdots, n$.

For almost all a and b,

$$(39)\qquad \Pr\{u(n; X_1, \cdots, X_n) \in S \mid t_0(n; X_1, \cdots, X_n) = a, \; s_0(n; X_1, \cdots, X_n) = b\}$$

is independent of (θ, σ) and equals

$$(40)\qquad \Pr\{u(n; Y_1, \cdots, Y_n) \in S \mid t_0(n, Y_1, \cdots, Y_n) = a, \; s_0(n; Y_1, \cdots, Y_n) = b\},$$

using notation indicated at the beginning of Sec. 1 On the other hand, from the hypotheses of the lemma, (39) also equals

$$(41)\qquad \Pr\left\{u(n; Y_1, \cdots, Y_n) \in S \mid t_0(n, Y_1, \cdots, Y_n) = \frac{a - \theta}{\sigma}, \; s_0(n; Y_1, \cdots, Y_n) = \frac{b}{\sigma}\right\}$$

From the equality of (40) and (41), it follows that the conditional distribution of u is independent of the conditioning values of t_0 and s_0, so that u is stochastically independent of (t_0, s_0). But t_0 and s_0 are mutually independent, since Corollary 2.2 applies. Hence the result.

This lemma can be used to compare the relative merits of s_0 and any other s asymptotically as $\sigma \to \infty$. Let us assume that the hypotheses of Lemma 4 are satisfied, that F is continuous and that t_0 also satisfies Assumptions III and V. Then we know that t_0 is the best statistic to use as "t", and both s_0 and s are eligible as the "s" statistic. Let

$$(42)\qquad J(u) = \Pr\{s(m; X_1, \cdots, X_m) \leqq u s_0(m; X_1, \cdots, X_m)\},$$

and $H(u)$, $H_0(u)$ denote the c.d.f.'s of s and s_0 respectively. It will be understood that we have the same m throughout the discussion. We already know

$$(43)\qquad \Pr\{t_0(n; X_1, \cdots, X_n) \leqq \theta + \sigma x\} = F\{x k_0(n)\}.$$

[2] My attention has been drawn to the fact that a general result of the type of those given in Lemmas 2 and 4 has been previously given, for boundedly complete sufficient statistics, by D. Basu [*Sankhya*, vol. 15 (1955), pp. 377–380.]

Let

(44) $$M(y) = \int_0^\infty F(yu)\, dH_0(u).$$

Then

(45) $$H(v) = \int_0^\infty H_0(v/u)\, dJ(u)$$

by Lemma 4, and we get

(46) $$\int_0^\infty F(yu)\, dH(u) = \int_0^\infty M(yu)\, dJ(u).$$

From (5), (8), and (9) on account of continuity, we have

(47) $$\rho = (\chi - \chi')/\delta, \qquad \rho_0 = (\chi_0 - \chi_0')/\delta,$$

and

(48) $$\begin{cases} M(\chi_0) = 1 - \alpha = \int_0^\infty M(\chi u)\, dJ(u), \\ M(\chi_0') = 1 - \beta = \int_0^\infty M(\chi' u)\, dJ(u). \end{cases}$$

Hence,

(49) $$(\chi_0 - \chi_0') = M^{-1}\left\{\int_0^\infty M(\chi u)\, dJ(u)\right\} - M^{-1}\left\{\int_0^\infty M(\chi' u)\, dJ(u)\right\}.$$

Now, from (14) and (15) we know that $E(N)$, $E(N_0) \to \infty$ as $\sigma \to \infty$, and

$$\frac{E(N_0)}{E(N)} \cong \frac{\int_0^\infty k_0^{-1}(\sigma\rho_0 u)\, dH_0(u)}{\int_0^\infty k^{-1}(\sigma\rho u)\, dH(u)}.$$

Suppose

(50) $$k(u) = u^{1/c},$$

where c is a constant $\geqq 1$. (This is the case in the normal and exponential populations.) Then

(51) $$\frac{E(N_0)}{E(N)} \cong \frac{\rho_0^c \int_0^\infty u^c\, dH_0(u)}{\rho^c \int_0^\infty u^c\, dH(u)} = \frac{\rho_0^c}{\rho^c \int_0^\infty u^c\, dJ(u)},$$

by (45).

Therefore

$$\rho\left\{\int_0^\infty u^c\, dJ(u)\right\}^{1/c} > \rho_0$$

implies that, asymptotically as $\sigma \to \infty$, $E(N_0) < E(N)$. However, since

$$\left\{\int_0^\infty u^c\, dJ(u)\right\}^{1/c} > \int_0^\infty u\, dJ(u),$$

if we can show that

$$(52) \qquad (\chi - \chi') \int_0^\infty u\, dJ(u) > \chi_0 - \chi_0',$$

this implies that asymptotically (t_0, s_0) is the best, or minimum-expected-sample-size, pair among those satisfying the initial assumptions.

We shall now prove (52) to hold in the two cases that matter. As previously noted, s_0 and as_0, where a is a constant >0, are equivalent statistics for our purpose, and hence in what follows we shall only consider as alternative candidates, statistics s which are not constant multiples of s_0, in other words, we assume that $J(u)$ has at least two points of increase.

EXAMPLE 1. Let $F(x) = \int_{-\infty}^{x} e^{-(u^2/2)}\, du/\sqrt{2\pi}$, and assume $\alpha < 0.5 < \beta$. $M(y)$ is Student's distribution, so that $M(0) = 0.5$. Hence

$$(53) \qquad \chi_0' < 0 < \chi_0 \quad \text{and} \quad \chi' < 0 < \chi.$$

Further, $M(y)$ is concave or convex according as $y > 0$ or <0, and therefore $M^{-1} \int_0^\infty M(yu)\, dJ(u) \lessgtr y \int_0^\infty u\, dJ(u)$ according as $y \gtrless 0$. From (53) and (49), (52) follows.

EXAMPLE 2. Let $F(x)$ be given by (17). Then

$$M(y) = \begin{cases} 0 & \text{if } \; y \leqq 0 \\ 1 - (1 + y)^{-\mu} & \text{if } \; y > 0, \end{cases} \qquad \text{where } \mu = m - 1 \geqq 1.$$

Consequently, all χ's are positive. Now let

$$(54) \qquad \begin{aligned} f(y) &= y \int_0^\infty u\, dJ(u) - M^{-1} \int_0^\infty M(yu)\, dJ(u) \\ &= y \int_0^\infty u\, dJ(u) - \left\{\int_0^\infty (1 + yu)^{-\mu}\, dJ(u)\right\}^{-1/\mu} - 1. \end{aligned}$$

Then

$$f'(y) = \int_0^\infty u\, dJ(u) - \int_0^\infty u(1 + yu)^{-\mu-1}\, dJ(u) \cdot \left\{\int_0^\infty (1 + yu)^{-\mu}\, dJ(u)\right\}^{-(\mu+1)/\mu}.$$

It is easily seen that $f'(y) > 0$ for $y > 0$; because the sign of $f'(y)$ is the same as that of

$$\left\{\int_0^\infty (1 + yu)^{-\mu}\, dJ(u)\right\}^{(\mu+1)/\mu} \int_0^\infty u\, dJ(u) - \int_0^\infty u(1 + yu)^{-\mu-1}\, dJ(u)$$

which is

$$>\left\{\int (1 + yu)^{-\mu}\, dJ(u)\right\}\left\{\int (1 + yu)^{-1}\, dJ(u)\right\} \int u dJ(u)$$

$$- \int u(1 + yu)^{-\mu-1}\, dJ(u)$$

$$> \int (1 + yu)^{-\mu}\, dJ(u) \int u(1 + yu)^{-1}\, dJ(u) - \int u(1 + yu)^{-\mu-1}\, dJ(u)$$

$$> 0,$$

since u and $(1 + yu)^{-1}$ are monotone in opposite directions for $y > 0$, and the same is true of $u(1 + yu)^{-1}$ and $(1 + yu)^{-\mu}$. Consequently, $f(y)$ is an increasing function of $y > 0$; (52) follows from (49), and asymptotically as $\sigma \to \infty$, (t_0, s_0) is the best pair of statistics to use.

REFERENCES

[1] C. Stein, "A two-sample test of a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 243–258.
[2] S. G. Ghurye and Herbert Robbins, "Two-stage procedures for estimating the difference between means," *Biometrika*, Vol. 41 (1954), pp. 146–152.
[3] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1939), pp. 391–421.

SAMPLING VARIANCES OF ESTIMATES OF COMPONENTS OF VARIANCE

By S. R. Searle

New Zealand Dairy Board, Wellington, N. Z.[1]

1. Outline. In earlier work (4) matrix methods have been developed for obtaining the sampling variances of estimates of components of variance. These rely on the fact that if $y = \mathbf{x}'F\mathbf{x}$ is a function of variables $\mathbf{x}$, having a multinormal distribution with variance-covariance matrix V, then the variance of y is given by

$$\text{var}(y) = 2 \text{ tr } (VF)^2. \tag{1}$$

The use of the method was demonstrated by obtaining for the case of a 1-way classification with unequal numbers in the sub-classes, the sampling variances of the estimates of variance components, as summarized in (1); it was then extended to the sampling variances of estimates of components of covariance.

The present paper makes further use of this matrix technique to obtain the sampling variances of estimates of components of variance from data in a 2-way classification having unequal sub-class numbers. The model assumed is Eisenhart's Model II, [2], and the method of estimating the components is taken to be Henderson's Method 1, [3]

2. Model and analysis of variance. The observations x_{ijk} are taken as having the linear model

$$x_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \epsilon_{ijk},$$

with $k = 1 \cdots n_{ij}$, $i = 1 \cdots a$, and $j = 1 \cdots b$ μ is a general mean, A_i and B_j are main effects, $(AB)_{ij}$ is an interaction and ϵ_{ijk} is residual error. Under the assumptions of the model, all terms (except μ) are taken as being normally distributed, with zero means, and variances σ_a^2, σ_b^2, σ_{ab}^2, and σ_ϵ^2, which we will write as α, β, γ and ϵ respectively.

For a sample of N observations in N' cells of this 2-way classification an analysis of variance can be written as

Term	d-f	Sums of Square
Between A classes	$a - 1$	$T_a - T_f = S_a$
Between B classes	$b - 1$	$T_b - T_f = S_b$
Interaction $A \times B$	$N' - a - b + 1$	$T_{ab} - T_a - T_b + T_f = S_{ab}$
Residual...	$N - N'$	$T_0 - T_{ab} = S_w$
Total	$N - 1$	$T_0 - T_f$

where the T's are uncorrected sums of squares. With $n_{i.} = \sum_j n_{ij}$, and $n_{.j}$

Received December 3, 1956; revised September 5, 1957.

[1] The author's present address is Animal Husbandry Dept., Cornell University, Ithaca, New York.

$= \sum_i n_{ij}$, and using customary notation for means,

$$T_a = \sum_i n_{i.} \bar{x}_{i..}^2, \qquad T_b = \sum_j n_{.j} \bar{x}_{.j.}^2,$$

$$T_{ab} = \sum_i \sum_j n_{ij} \bar{x}_{ij.}^2, \quad T_f = N\bar{x}_{...}^2, \qquad \text{and } T_0 = \sum_i \sum_j \sum_k x_{ijk}^2 .$$

We may note in passing that not all the expressions in the "sums of squares" column are in fact sums of squares, notably the interaction term. It would be more correct to label this column "quadratic forms" but the terminology "sums of squares" has historical precedence and will be retained.

Henderson's first method [3] for estimating the components of variance is to equate each of the first four lines in the above analysis to its expected value. Denoting the resulting estimates of α, β, γ, and ϵ as $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\epsilon}$, the equations for obtaining them are

$$(2) \quad \begin{cases} T_a - T_f = S_a = (N - k_1)\hat{\alpha} + (k_{12} - k_2)\hat{\beta} \\ \qquad + (k_{12} - k_3)\hat{\gamma} + (a - 1)\hat{\epsilon} \\ T_b - T_f = S_b = (k_{21} - k_1)\hat{\alpha} + (N - k_2)\hat{\beta} \\ \qquad + (k_{21} - k_3)\hat{\gamma} + (b - 1)\hat{\epsilon} \\ T_{ab} - T_a - T_b + T_f = S_{ab} = (k_1 - k_{21})\hat{\alpha} + (k_2 - k_{12})\hat{\beta} \\ \qquad + (N - k_{12} - k_{21} + k_3)\hat{\gamma} \\ \qquad + (N' - a - b + 1)\hat{\epsilon} \end{cases}$$

$$(3) \quad T_0 - T_{ab} = S_w = (N - N')\hat{\epsilon}$$

where the k's are functions of the n_{ij}'s, namely

$$k_{12} = \sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} \qquad k_{21} = \sum_j \frac{\sum_i n_{ij}^2}{n_{.j}},$$

$$k_1 = \frac{1}{N} \sum_i n_{i.}^2, \qquad k_2 = \frac{1}{N} \sum n_{.j}^2, \qquad \text{and } k_3 = \frac{1}{N} \sum n_{ij}^2 .$$

3. Variances required. In the analysis of variance S_w has a χ^2-distribution with $N - N'$ degrees of freedom. Hence, from Eq. (3) the variance of $\hat{\epsilon}$ is

$$\sigma_{\hat{\epsilon}}^2 = \frac{2\epsilon^2}{N - N'} .$$

Using (3), Eqs. (2) give $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ as linear functions of S_a, S_b, S_{ab}, and $\hat{\epsilon}$. But S_w, and hence $\hat{\epsilon}$, is distributed independently of S_a, S_b, and S_{ab}. Hence the variances and covariances of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ can be obtained as linear functions of $\sigma_{\hat{\epsilon}}^2$ and the variances and covariances of S_a, S_b, and S_{ab}. By the nature of the S's it is easier to consider the variances and covariances of T_a, T_b, T_{ab}, and T_f.

Writing P for the matrix of coefficients of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ in Eqs. (2) these equations can be written as

$$\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} T_a \\ T_b \\ T_{ab} \\ T_f \end{pmatrix} = P \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} + \hat{\epsilon} \begin{pmatrix} a - 1 \\ b - 1 \\ N' - a - b + 1 \end{pmatrix},$$

which we may write as

$$H\mathbf{t} = P\mathbf{v} + \hat{\epsilon}\mathbf{m}.$$

Since $\hat{\epsilon}$ is independent of the terms in $H\mathbf{t}$, the variance-covariance matrix of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$, var $(\mathbf{vv}')$, can be expressed in terms of the variance-covariance matrix of the T's, var $(\mathbf{tt}')$, as

(4) $$\text{var}\,(\mathbf{vv}') = P^{-1}[H \,\text{var}\,(\mathbf{tt}')H' + \mathbf{mm}'\sigma_{\hat{\epsilon}}^2]P^{-1'}$$

and

$$\text{cov}\,(\hat{\epsilon}\mathbf{v}) = -P^{-1}\mathbf{m}\sigma_{\hat{\epsilon}}^2 .$$

The unknown term in these expressions is var $(\mathbf{tt}')$ the variance-covariance matrix of T_a, T_b, T_{ab}, and T_f, which we now proceed to obtain, term by term.

4. Matrix definitions and expressions. Let U be a matrix having a one for every element, its order being denoted by subscripts, thus:—

U-matrix (*all elements 1*)	Order
$U_{ij,kl}$	$n_{ij} \times n_{kl}$
$U_{ij,k}$	$n_{ij} \times n_{k}$
U_{ij}	$n_{ij} \times n_{ij}$
U_{i}	$n_{i} \times n_{i}$
U_{N}	$N \times N$

Define W-matrices in terms of the U's:

$$W_{ij} = \frac{1}{n_{ij}} U_{ij} ,$$

$$W_{i.} = \frac{1}{n_{i.}} U_{i.} ,$$

$$W_{.j} = \frac{1}{n_{.j}} U_{.j}, \text{ and } W_N = \frac{1}{N} U_N .$$

Then C-matrices are defined, of order $N \times N$, whose only non-zero submatrices are W's along the diagonal:

C_a has	$W_{i.}(i = 1 \cdots a)$	in the diagonal,
C_b has	$W_{.j}(j = 1 \cdots b)$	in the diagonal,
C_{ab} has	$W_{ij}(i = 1 \cdots a, j = 1 \cdots b)$,	in the diagonal.

Finally we define D-matrices, the same as C-matrices only having U-matrices instead of W-matrices in their diagonals.

Let x′ be the row vector of the N x_{ijk}'s, arrayed in order, $k = 1 \cdots n_{ij}$, within j-classes, within each i-class; i.e.,

$$\mathbf{x}' = (x_{111} \cdots x_{11n_{11}} \quad x_{121} \cdots x_{12n_{12}} \cdots x_{ab1} \cdots x_{abn_{ab}}).$$

Then if w′ is the vector of the x's arrayed in k-order within i-classes within each j-class, w′ will be a transform of x′, $\mathbf{w}' = \mathbf{x}'R'$, say, where R is an orthogonal elementary operational matrix of order N, of identity matrices I.

The T's can now be expressed in terms of these vectors and matrices:

$$T_a = \mathbf{x}'C_a\mathbf{x},$$

$$T_b = \mathbf{w}'C_b\mathbf{w} = \mathbf{x}'R'C_bR\mathbf{x} = \mathbf{x}'B\mathbf{x}\text{ , say.}$$

$$T_{ab} = \mathbf{x}'C_{ab}\mathbf{x} = \mathbf{w}'C_{ba}\mathbf{w},$$

$$T_f = \mathbf{x}'U_N\mathbf{x}.$$

In C_{ab} the W_{ij} in the diagonal are in j-order within i-order; in C_{ba} they are in i-order within j-order.

V, the variance-covariance matrix of the x_{ijk}'s appropriate to x′ can be written as

$$V = J + K,$$

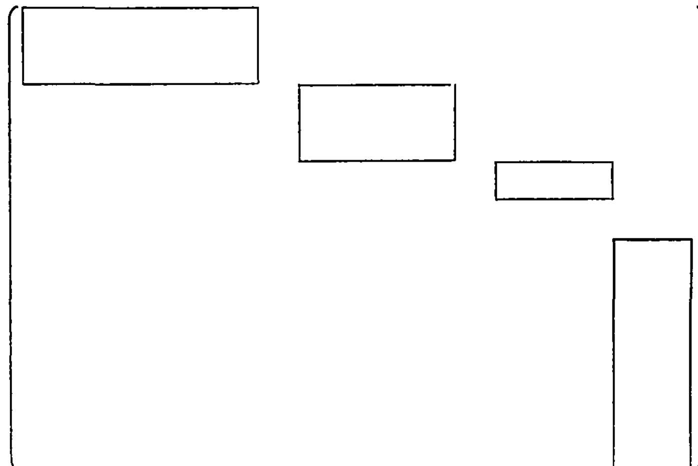
where

$$J = \alpha D_a + (\beta + \gamma)D_{ab} + \epsilon I,$$

and

$$K = \begin{pmatrix} 0 & K_{12} & K_{13} \cdots K_{1a} \\ K_{21} & 0 & \\ \vdots & & \\ K_{a1} & & 0 \end{pmatrix}$$

with $K_{ii'}$, $i \neq i'$, $i, i' = 1 \cdots a$, of order $n_{i.} \times n_{i'.}$ has all elements zero except those in b rectangular matrices $\beta U_{ij,i'j}$, $j = 1 \cdots b$. These b matrices lie "corner to corner" across $K_{ii'}$ thus:



For example the V matrix for a sample of 7 observations with $n_{11} = 1$, $n_{12} = 2$, $n_{21} = 3$ and $n_{22} = 1$ would be

$$\left(\begin{array}{ccc|cccc}
\alpha+\beta+\gamma+\epsilon & \alpha & \alpha & \beta & \beta & \beta & \cdot \\
\alpha & \alpha+\beta+\gamma+\epsilon & \alpha+\beta+\gamma & & \cdot & & \beta \\
\alpha & \alpha+\beta+\gamma & \alpha+\beta+\gamma+\epsilon & & & & \beta \\
\hline
\beta & \cdot & \cdot & \alpha+\beta+\gamma+\epsilon & \alpha+\beta+\gamma & \alpha+\beta+\gamma & \alpha \\
\beta & & \cdot & \alpha+\beta+\gamma & \alpha+\beta+\gamma+\epsilon & \alpha+\beta+\gamma & \alpha \\
\beta & \cdot & \cdot & \alpha+\beta+\gamma & \alpha+\beta+\gamma & \alpha+\beta+\gamma+\epsilon & \alpha \\
\cdot & \beta & \beta & \alpha & \alpha & \alpha & \alpha+\beta+\gamma+\epsilon
\end{array}\right)$$

This is the variance-covariance matrix appropriate to **x**. that for **w** will be RVR'.

5. Variances and covariances. T_a, T_b, T_{ab}, and T_f have now been expressed *in the form* $\mathbf{x}'F\mathbf{x}$, *and the variance-covariance matrix has also been obtained* The sampling variances of the T's will be found from (1), by evaluating $2 \text{ trace } (VF)^2$ for each of them.

5.1. $$\text{var}(T_a) = 2 \text{ tr } (VC_a)^2.$$

VC_a can be expressed as

$$(VC_a) = \begin{pmatrix} P_{11} P_{12} \cdot & \cdot P_{1a} \\ P_{21} \,. & \\ \vdots & \cdot \\ P_{a1} & P_{aa} \end{pmatrix},$$

where P_{ii} is a column of matrices $x_{ij}U_{ij,i}$, $(j = 1 \cdot \cdot b)$ with

$$x_{ij} = (1/n_{i.})(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon).$$

Similarly $P_{ii'}$ is a column of matrices $w_{ij}U_{ij,i'}$, $(j = 1 \cdot\cdot b)$, with

$$w_{i'j} = (n_{i'j}/n_{i'.})\beta.$$

VC_a has here been partitioned into P-matrices, which themselves have been partitioned into sub-matrices of the U-type. Trace $(VC_a)^2$ will therefore depend on two properties of these U matrices, that

(5) $$U_{ij,pq}U_{pq,rs} = n_{pq}U_{ij,rs}$$

and

$$\text{tr } (U_{ij}) = n_{ij}$$

Hence

(6) $$\text{tr } (U_{ij,pq}U_{pq,ij}) = n_{ij}n_{pq}\,.$$

Using these results we have

$$\operatorname{tr}(VC_a)^2 = \sum_i \sum_{i'} \operatorname{tr}(P_{ii'}P_{i'i})$$

$$= \sum_i \operatorname{tr}(P_{ii}^2) + \sum_i \sum_{i' \neq i} \operatorname{tr}(P_{ii'}P_{i'i})$$

$$= \sum_i \sum_j n_{ij}x_{ij} \sum_j n_{ij}x_{ij} + \sum_i \sum_{i' \neq i} \sum_j n_{ij}w_{i'j} \sum_j n_{i'j}w_{ij}.$$

On substituting for x_{ij} and w_{ij} this gives

$$\tfrac{1}{2}\operatorname{var}(T_a) = \sum_i [\sum_j n_{jj}(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)/n_{i.}]^2 + \sum_i \sum_{i' \neq i} \frac{(\sum_j n_{ij}n_{i'j})^2}{n_{i.}n_{i'.}}\beta^2.$$

5.2. $$\operatorname{var}(T_{ab}) = 2\operatorname{tr}(VC_{ab})^2.$$

V and C_{ab} are such that their product can be written as

$$VC_{ab} = L + K,$$

where K is as in V, and

$$L = \alpha D_a + (\beta + \gamma)D_{ab} + \epsilon C_{ab}.$$

Hence,

(7) $$VC_{ab} = V + \epsilon(C_{ab} - I).$$

Since V and C_{ab} are symmetric, VC_{ab} is also, and hence squaring (7) gives

$$(VC_{ab})^2 = V^2 + \epsilon^2(C_{ab}^2 - I).$$

Hence,

$$\tfrac{1}{2}\operatorname{var}(T_{ab}) = \operatorname{tr} V^2 + \epsilon^2(\operatorname{tr} C_{ab}^2 - \operatorname{tr} I)$$

$$= \sum_i \sum_j n_{ij}[(\alpha + \beta + \gamma + \epsilon)^2 + (n_{ij} - 1)(\alpha + \beta + \gamma)^2 + (n_{i.} - n_{ij})\alpha^2 + (n_{.j} - n_{ij})\beta^2] + \epsilon^2\left[\sum_i \sum_j n_{ij}n_{ij}\frac{1}{n_{ij}^2} - N\right],$$

which reduces to

$$\tfrac{1}{2}\operatorname{var}(\mathrm{T}_{ab}) = \sum_i \sum_j n_{ij}[n_{ij}(\alpha + \beta + \gamma + \epsilon/n_{ij})^2 + (n_{i.} - n_{ij})\alpha^2 + (n_{.j} - n_{ij})\beta^2].$$

5.3. $$\operatorname{var}(T_f) = 2\operatorname{tr}(VW_N)^2.$$

Similar to the form of the P-matrices in 5.2, VW_N can be expressed as a column of matrices $y_{ij}U_{ij.N}$, $(i = 1 \cdots a, j = 1 \cdots b)$, where

$$y_{ij} = (n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/N.$$

Hence,

$$\text{tr}\,(VW_N)^2 = \sum_i \sum_j n_{ij} y_{ij} \sum_i \sum_j n_{ij} y_{ij},$$

giving

$$\tfrac{1}{2}\,\text{var}\,(T_f) = [\sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon(/N]^2.$$

5.4. In general, for any two square matrices of the same order, A and B say, it can be shown that $\text{tr}\,(A + B)^2 = \text{tr}\,A^2 + \text{tr}\,B^2 + 2\,\text{tr}\,AB$. If then, $\hat{a}$ and $\hat{b}$ are two function of the same set of variables such that $\text{var}\,(\hat{a}) = 2\,\text{tr}\,A^2$, and $\text{var}\,(\hat{b}) = 2\,\text{tr}\,B^2$, it follows at once that

$$\text{(8)} \qquad \text{cov}\,(\hat{a}\hat{b}) = 2\,\text{tr}\,AB = 2\,\text{tr}\,BA.$$

This result will be used for obtaining the covariances among T_a, T_b, T_{ab}, and T_f.

5.5.

$$\text{cov}\,(T_a, T_{ab}) = 2\,\text{tr}\,(VC_a)(VC_{ab}).$$

In 5.1 VC_a has been partitioned into P_{ii}'s and $P_{ii'}$'s. If VC_{ab}, expressed as $L + K$ in 5.2 is partitioned in the same manner, into L_{ii}'s and $K_{ii'}$'s, then

$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_{ab}) = \sum_i \sum_{l=1}^{n_i} (\text{inner product of } l\text{'th row of } P_{ii} \text{ and } l\text{ th column of } L_{ii})$$
$$+ \sum_i \sum_{i' \neq i} \sum_{l=1} (\text{inner product of } l\text{'th row of } P_{ii'} \text{ and } l\text{ th column of } K_{i'i})$$

and after substitution this reduces to

$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_{ab}) = \sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)^2/n_i$$
$$+ \beta^2 \sum_i \sum_j n_{ij}^2(n_{.j} - n_{ij})/n_{i.}.$$

5.6.

$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_f) = \text{tr}\,(VW_N)(VC_a).$$

Using 5.1 and 5.4, and Eq. (6), this can be expressed as

$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_f) = \sum_i \sum_j n_{ij} y_{ij} \left(\sum_j n_{ij} x_{ij} + \sum_j \sum_{i' \neq i} n_{i'j} w_{ij}\right).$$

which on substitution for the x's, y's and w's, reduces to

$$\tfrac{1}{2}\,\text{cov}\,(T_a, T_f) = \sum_i \sum_j \frac{n_{ij}}{N}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon) \cdot \left(n_{i.}\alpha + \beta\frac{\sum_j n_{.j} n_{ij}}{n_{i.}} + \gamma\frac{\sum_j n_{ij}^2}{n_{i.}} + \epsilon\right).$$

5.7.

$$\tfrac{1}{2}\,\text{cov}\,(T_{ab}, T_f) = \text{tr}\,(VW_N)(VC_{ab})$$
$$= \sum_i \sum_j n_{ij} y_{ij} \left[\sum \text{terms in } ij\text{'th column of } V + \epsilon(C_{ab} - I)\right]$$
$$= \sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2/N.$$

5.8. In Eq. (8) it is required that $\hat{a}$ and $\hat{b}$ be functions of the same set of variables; therefore, in terms of paragraph 4, the covariance of T_a and T_b must be expressed as

$$\text{cov}\,(T_a\,,\,T_b) = 2\,\text{tr}\,(VC_a)(VB).$$

This covariance is a little more cumbersome to evaluate than previous ones; the method used is essentially a generalization of earlier paragraphs.

B is the same form as V, (4.5) but with matrices $(1/n_{.j})U_{ij}$ in the diagonal $j = 1 \cdots b$, for $i = 1 \cdots a$, and with $K_{ii'}$-matrices having terms

$$(1/n_{.j})U_{ij,i'j}\,.$$

Now partition V into matrices $(V)_{ij:kl}$ of order $n_{ij} \times n_{kl}$, there being four different forms of this matrix according as k and l are equal or not equal to i and j respectively, namely:

$$(V)_{ij:ij} = (\alpha + \beta + \gamma)U_{ij} + eI;$$

$$(V)_{ij:il} = \alpha U_{ij,il} \qquad \text{for } l \neq j;$$

$$(V)_{ij:kj} = \beta U_{ij,kj}\,, \qquad \text{for } k \neq l;$$

$$(V)_{ij:kl} = 0.U_{ij,kl} \text{ a zero matrix,} \qquad \text{for } k \neq i,\, l \neq j.$$

B can be partitioned similarly for $k \neq i$ and $l \neq j$:

$$(B)_{ij:ij} = \frac{1}{n_{.j}}\,U_{ij}\,,$$

$$(B)_{il:ij} = 0.U_{il,ij}\,,$$

$$(B)_{kj:ij} = \frac{1}{n_{:j}}\,U_{kj,ij}\,,$$

$$(B)_{kl:ij} = 0.U_{kl,ij}\,.$$

Consider now the identity

$$(9) \qquad (VB)_{pq:tu} = \sum_f \sum_g V_{pq:fg} B_{fg:tu}\,,$$

whose right-hand side can be expanded as

$$(V)_{pq:tu}(B)_{tu:tu} + \sum_{f \neq t} (V)_{pq:fu}(B)_{fu:tu} + \sum_{g \neq u} (V)_{pq:tg}(B)_{tg:tu} + \sum_{f \neq t} \sum_{g \neq u} (V)_{pq:fg}(B)_{fg:tu}$$

or as

$$(V)_{pq:pq}(B)_{pq:tu} + \sum_{f \neq p} (V)_{pq:fq}(B)_{fq:tu} + \sum_{g \neq q} (V)_{pq:pg}(B)_{pg:tu} + \sum_{f \neq p} \sum_{g \neq q} (V)_{pq:fg}(B)_{fg:tu}\,.$$

These expressions are true for any values of the subscripts.

Applying this identity to the partitioned forms of V and B given above, and using the principle of (5) in 5.1 gives

$$(VB)_{ij\cdot ij} = n_{ij}(\alpha + \beta + \gamma)/n_{.j}\,U_{ij} + \epsilon/n_{.j}\,U_{ij} + \beta \sum_{k \neq i} \frac{1}{n_{.j}}\, U_{ij,kj}\,U_{kj,ij}$$

$$= (n_{ij}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/n_{.j}\,U_{ij} = b_{ij}\,U_{ij}, \quad \text{say.}$$

Similarly, for $r \neq i$, and $s \neq j$,

$$(VB)_{is:ij} = \alpha \frac{n_{ij}}{n_{.j}}\, U_{is,ij} = b'_{ij}\, U_{is,ij}, \quad \text{say,}$$

$$(VB)_{rj\cdot ij} = (n_{ij}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/n_{.j}\,U_{rj,ij} = b_{ij}\,U_{rj,ij},$$

$$(VB)_{rs\cdot ij} = \alpha \frac{n_{rj}}{n_{.j}}\, U_{rs,ij} = b'_{rj}\, U_{rs,ij}.$$

Likewise:

$$(VC_a)_{ij.ij} = (n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)/n_{i.}\,U_{ij} = a_{ij}\,U_{ij}, \quad \text{say,}$$

$$(VC_a)_{ij.is} = (n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)/n_{i.}\,U_{ij,is} = a_{ij}\,U_{ij,is},$$

$$(VC_a)_{ij:rj} = \beta \frac{n_{rj}}{n_{r.}}\, U_{ij,rj} = a'_{rj}\, U_{ij,rj}, \quad \text{say,}$$

$$(VC_a)_{ij:rs} = \beta \frac{n_{rj}}{n_{r}}\, U_{ij,rs} = a'_{rj}\, U_{ij,rs}.$$

Now

$$\tfrac{1}{2}\operatorname{cov}(T_a T_b) = \operatorname{tr}(VC_a)(VB)$$

$$= \sum_i \sum_j \operatorname{tr}(VC_a \cdot VB)_{ij:ij}$$

$$= \sum_i \sum_j \Big[\operatorname{tr} \sum_r \sum_s (VC_a)_{ij:rs}(VB)_{rs:ij}\Big].$$

Applying the identity (9) and results (5) and (6) again, and using the forms of the elements of the sub-matrices of VC_a and VB given above, gives

$\tfrac{1}{2}\operatorname{cov}(T_a, T_b)$

$$= \sum_i \sum_j n_{ij}\Big\{n_{ij}a_{ij}b_{ij} + \sum_{s \neq j} n_{is}a_{ij}b'_{ij} + \sum_{r \neq i} n_{rj}a'_{rj}b_{rj} + \sum_{r \neq i}\sum_{s \neq j} n_{rs}a'_{rj}b'_{rj}\Big\}.$$

On substituting for the a's and b's, this reduces to

$$\tfrac{1}{2}\operatorname{cov}(T_a, T_b) = \sum_i \sum_j \frac{n_{ij}^2}{n_{i.}n_{.j}}\,(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2.$$

5.9. We have now found some of the variances and covariances of T_a, T_b, T_{ab} and T_f. These and those which follow from them by symmetry, are summarized in the following table.

VARIANCES AND COVARIANCES OF UNCORRECTED SUMS OF SQUARES

var (T_a)

$$= 2\left\{\sum_i [\sum_j n_{ij}(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)/n_{i.}]^2 + \sum_i \sum_{i' \neq i} \frac{(\sum_j n_{ij}n_{i'j})^2}{n_{i.}n_{i'.}}\beta^2\right\}$$

var (T_b)

$$= 2\left\{\sum_j [\sum_i n_{ij}(n_{ij}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/n_{.j}]^2 + \sum_j \sum_{j' \neq j} \frac{(\sum_i n_{ij}n_{ij'})^2}{n_{.j}n_{.j'}}\alpha^2\right\}$$

var (T_{ab})

$$= 2\sum_i \sum_j n_{ij}[n_{ij}(\alpha + \beta + \gamma + \epsilon/n_{ij})^2 + (n_{i.} - n_{ij})\alpha^2 + (n_{.j} - n_{ij})\beta^2]$$

var (T_f)

$$= 2[\sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)/N]^2$$

cov (T_a, T_b)

$$= 2\sum_i \sum_j \frac{n_{ij}^2}{n_{i.}n_{.j}}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2$$

cov (T_a, T_{ab})

$$= 2\{\sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{ij}\beta + n_{ij}\gamma + \epsilon)^2/n_{i.} + \beta^2 \sum_i \sum_j n_{ij}^2(n_{.j} - n_{ij})/n_{i.}\}$$

cov (T_a, T_f)

$$= 2\sum_i \sum_j \frac{n_{ij}}{N}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)\left(n_{i.}\alpha + \beta\frac{\sum_j n_{.j}n_{ij}}{n_{i.}} + \gamma\frac{\sum_j n_{ij}^2}{n_{i.}} + \epsilon\right)$$

cov (T_b, T_{ab})

$$= 2\{\sum_i \sum_j n_{ij}(n_{ij}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2/n_{.j} + \alpha^2 \sum_i \sum_j n_{ij}^2(n_{i.} - n_{ij})/n_{.j}\}$$

cov (T_b, T_f)

$$= 2\sum_i \sum_j \frac{n_{ij}}{N}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)\left(\alpha\frac{\sum_i n_{i.}n_{ij}}{n_{.j}} + \beta n_{.j} + \gamma\frac{\sum_i n_{ij}^2}{n_{.j}} + \epsilon\right)$$

cov (T_{ab}, T_f)

$$= 2\sum_i \sum_j n_{ij}(n_{i.}\alpha + n_{.j}\beta + n_{ij}\gamma + \epsilon)^2/N$$

The expressions in the above table are those of the elements of the matrix var (tt′) of Eq. (4). These elements are quadratic functions of the variance

components α, β, γ, and ϵ, with coefficients being sums of functions of the n_{ij}'s. The other terms in (4) are not such as would simplify var (vv') if the elements of var (tt') as now known were inserted into (4), and therefore, as in any numerical case after calculating the expressions in the table these steps will be quite straightforward, it seems convenient to leave the results in their present form.

6. Balanced Data. It is easily shown that the formulae developed in the last paragraph reduce to the well-known results for balanced data when all the n_{ij} are put equal to n. For example, consider the variance of S_a. From the Analysis of Variance table, the expected value of S_a is given by

$$E(S_a) = (a - 1)(bn\alpha + n\gamma + \epsilon).$$

Then

$$\text{var}\,(T_a) = 2[a(bn\alpha + n\beta + n\gamma + \epsilon)^2 + a(a - 1)n^2\beta^2],$$

$$\text{var}\,(T_f) = 2(bn\alpha + an\beta + n\gamma + \epsilon)^2,$$

$$\text{cov}\,(T_a, T_f) = 2(bn\alpha + an\beta + n\gamma + \epsilon)^2.$$

Hence,

$$\begin{aligned} \text{var}\,(S_a) &= \text{var}\,(T_a - T_f) \\ &= 2(a - 1)(bn\alpha + n\gamma + \epsilon)^2 \\ &= 2[E(Sa)]^2/(a - 1) \end{aligned}$$

and with $M_a = S_a/(a - 1)$, this gives the familiar result for mean squares

$$\text{var}\,(M_a) = 2[E(M_a)]^2/(a - 1).$$

Results similar to this can be obtained for M_b and M_{ab}, the mean squares [illegible] B-effects and interaction.

7. Conclusion. Matrix methods have been developed for finding the [illegible] variances of estimates of components of variance. In earlier work (4) [illegible] used for data in a 1-way classification, and this paper has extended [illegible] data for a 2-way classification, with unequal numbers of observations [illegible] classes. The estimates of the components of variance for main effects [illegible] tion are expressed as linear functions of the corrected sums of [illegible] estimate of the error variance component. By expressing the [illegible] squares as functions of the uncorrected sums of squares, the [illegible] matrix of the estimates of the components of variance has [illegible] function of that for the uncorrected sums of squares, (Eq [illegible] then been found for the elements of this, the variance-[illegible] uncorrected sums of squares. It has been checked that [illegible] sumed balanced, i.e., all n_{ij} equal to n, these expressions [illegible] priate forms for variances of mean squares then [illegible]

tributions. Estimates with any optimum properties have not been obtained, and it would seem that the only feasible estimation procedure in a practical case would be that of replacing the variance components in these formulae by their estimates.

It is hoped that these methods can next be extended to data in a 3-way classification with unequal subclass numbers, still based on Eisenhart's Model II and using Henderson's Method I for estimation.

REFERENCES

[1] S. L. Crump, "Present status of variance components analysis," *Biometrics*, Vol. 7 (1951), pp. 1–16.

[2] C. Eisenhart, "Assumptions underlying analysis of variance," *Biometrics*, Vol. 3 (1947), pp. 1–21.

[3] C. R. Henderson, "Estimation of variance and covariance components," *Biometrics*, Vol. 9 (1953), pp. 226–252.

[4] S. R. Searle, "Matrix methods in components of variance and covariance analysis," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 737–748.

ON SOME DISTRIBUTIONS RELATED TO THE STATISTIC D_n^+

By Z. W. Birnbaum and Ronald Pyke[1]

University of Washington

1. Introduction and summary. Let $X_1 < X_2 < \cdots < X_n$ be a sample of size n, ordered increasingly, of a one-dimensional random variable X which has the continuous cumulative distribution function F. It is well known, [1], that the statistic

(1) $$D_n^+ = \sup_{-\infty < x < +\infty} \{F_n(x) - F(x)\},$$

where $F_n(x)$ is the empirical distribution function determined by $X_1, X_2, \cdots, X_n$, has a probability distribution independent of F. One may, therefore, assume that X has the uniform distribution in (0, 1) and, observing that the supremum in (1) must be attained at one of the sample points, write without loss of generality

(2) $$D_n^+ = \max_{1 \leq i \leq n} (i/n - U_i),$$

where $U_1 < U_2 < \cdots < U_n$ is an ordered sample of a random variable with uniform distribution in (0, 1).

For a given $n > 0$ define the random variable i^* as that value of i, determined uniquely with probability 1, for which the maximum in (2) is reached, i.e., such that

(3) $$D_n^+ = i^*/n - U_{i^*},$$

and write

(3.1) $$U_{i^*} = U^*.$$

The main object of this paper is to obtain the distribution functions of (i^*, U^*), of i^* and of U^*. The asymptotic distribution of $\alpha_n = i^*/n$ is also investigated, and bounds are obtained on the difference between the exact and the asymptotic distribution.

A number of general identities, which are not commonly known, have been verified and used in proving the above-mentioned results. Since these identities may be helpful in other problems of this type, they are separated from the main proofs and appear in the next section.

Received March 25, 1957; revised July 26, 1957.

[1] Research under the sponsorship of the Office of Naval Research. The second author's research was also supported by the Ontario Research Foundation. This paper was presented at the Seattle meeting of the Institute in August, 1956

2. Some useful lemmas.

LEMMA 1. *For all real a, b and integer $n \geqq 0$*

$$\sum_{i=0}^{n} \binom{n}{i} (a+i)^i (b-i)^{n-1} = n! \sum_{i=0}^{n} \frac{(a+b)^i}{i!}. \tag{4}$$

PROOF. The identity

$$(b-n) \sum_{i=0}^{n} \binom{n}{i} (a+i)^i (b-i)^{n-i-1} = (a+b)^n, \tag{5}$$

for all real a, b and integer $n \geqq 0$ (for $b = n$ the left-hand term is defined as the limit for $b \to n$) was proven by Abel ([2], Vol. 1, p. 102). Denoting the left side of (4) by $f_n(a, b)$ we have

$$f_n(a,b) - nf_{n-1}(a,b) = (b-n) \sum_{i=0}^{n} \binom{n}{i} (a+i)^i (b-i)^{n-i-1} = (a+b)^n$$

by (5). For $n = 1$, (4) is obviously true. Assuming it is true for $n - 1$ we have

$$f_n(a,b) = nf_{n-1}(a,b) + (a+b)^n = n! \sum_{i=0}^{n} \frac{(a+b)^i}{i!},$$

which completes the proof of (4) by induction.

LEMMA 2. *For all real a, b and integers $n \geqq 0$*

$$\sum_{i=0}^{n-1} \binom{n}{i} (a+i)^i (b-i)^{n-i-1} = \sum_{i=0}^{n-1} (a+b)^i (a+n)^{n-i-1}. \tag{6}$$

PROOF. For $b \neq n$, the left side of (6) is by Abel's identity (5) equal to

$$[(a+b)^n - (a+n)^n](b-n)^{-1},$$

which is equal to the right-hand side of (6), summed as a geometric progression. That (6) is true for $b = n$ follows from the continuity of both sides of (6).

LEMMA 3. *For all real a, b and integers $n > 0$*

$$\begin{aligned}(a-1)(b-n) &\sum_{i=0}^{n} \frac{1}{i+1} \binom{n}{i} (a+i)^i (b-i)^{n-i-1} \\ &= \frac{1}{n+1} [(a+b)^n (a+b-n-1) - (b+1)^n (b-n)].\end{aligned} \tag{7}$$

PROOF. Since $(a-1)/(i+1) = (a+i)/(i+1) - 1$, we may write

$$\begin{aligned}(a-1)(b-n) &\sum_{i=0}^{n} \frac{1}{i+1} \binom{n}{i} (a+i)(b-i)^{n-i-1} \\ &= \frac{b-n}{n+1} \sum_{i=0}^{n} \binom{n+1}{i+1} (a-1+i+1)^{i+1} (b+1-i-1)^{n+1-(i+1)-1} \\ &\quad - (b-n) \sum_{i=0}^{n} \binom{n}{i} (a+i)^i (b-i)^{n-i-1}.\end{aligned}$$

Applying Lemma 2 to the first sum and identity (5) to the second, one concludes that this is equal to the right-hand side of (7).

COROLLARY 1. *For all integers $n > 0$ we have*

$$\sum_{i=0}^{n-1} \frac{1}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1} = (n+1)^{n-1}. \tag{8}$$

PROOF. For $a = 0$, (7) yields

$$\sum_{i=0}^{n-1} \frac{1}{i+1} \binom{n}{i} i^i (b-i)^{n-i-1} = \frac{1}{n+1}\left[\frac{n^n - b^n}{n-b} - b^n + (b+1)^n\right]$$

and (8) follows for $b \to n$.

3. The distributions of (i^*, U^*), i^* and U^*. The following notations will be used: for any function f, denote by $[f \varepsilon B]$ that subset of the domain of f on which f takes values in B, a subset of the range of f; for any univariate distribution function, F, let P_F denote the n-dimensional product measure determined by the probability measure associated with F, without a subscript, P will be that measure determined by the uniform distribution function; the value of n, though suppressed in the notation, shall always be made clear by the particular circumstances of its use; furthermore, for $j = 1, 2, \cdots n$ and $u \varepsilon [0, 1]$, set

$$p_j = P[i^* = j], \qquad G^*(u, j) = P[U^* \leqq u, i^* \leqq j], \qquad H^*(u) = P[U^* \leqq u]; \tag{9}$$

for real x, $[x]$ denotes the greatest integer less than x.

All the theorems of this section are stated at the outset, and the proofs are then presented in what appears a natural sequence.

THEOREM 1. *The probabilities for i^* are given by*

$$p_j = n^{-n} \sum_{i=n-j}^{n-1} \frac{1}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1} \tag{10}$$

for $j = 1, 2, \cdots n$.

THEOREM 2. *The joint probability distribution of i^* and U^* is given for all $u \varepsilon [0, 1]$ by*

$$G^*(u, k) = \sum_{j=1}^{k} K(u, j) \qquad (k = 1, 2, \cdots n),$$

where

$$K(u, j) = P[U^* \leqq u, i^* = j] =$$

$$\begin{cases} p_j & \text{if } nu \geqq j \\ n^{-1} \sum_{i=j}^{n} \binom{n}{i} (i-u)^{n-i-1} (i - nu) n^{-i} \sum_{t=[nu]}^{i} \binom{i}{t} (nu - t - 1)^{i-t} (t+1)^{t-1} & \\ & \text{if } nu < j. \end{cases} \tag{11}$$

THEOREM 3. *The random variable U^* is uniformly distributed over* [0, 1].

PROOF OF THEOREM 2. For $j = 1, 2, \cdots, n$, consider the events

$$B_j = [U^* < u; i^* = j] = [U_j < u; j/n - U_j > i/n - U_i, (i \neq j)].$$

Employing the transformation

$$Z_j = U_j; \qquad Z_i = U_i - U_j, \qquad (i \neq j)$$

one obtains

$$B_j = [Z_j \leqq u; Z_i > (i - j) / n, (i \neq j)].$$

Setting $Z = (Z_1, Z_2, \cdots Z_n)$, the joint probability element of Z for j fixed is

$$dH_j(Z) = n!\, dZ_1\, dZ_2 \cdots dZ_n$$

for

$$[-Z_j \leqq Z_1 < Z_2 \leqq \cdots \leqq Z_{j-1} \leqq 0 \leqq Z_{j+1} \leqq \cdots \leqq Z_n \leqq 1 - Z_j]$$

and zero elsewhere. Assume u and j fixed such that $nu < j$ and $1 < j < n$. Writing $\lambda = [nu]$, one has

$$K(u, j) = \int_{B_j} dH_j(Z)$$
$$= n! \int_{-1/n}^{0} dZ_{j-1} \int_{-2/n}^{Z_{j-1}} dZ_{j-2} \cdots \int_{-\lambda/n}^{Z_{j-\lambda+1}} dZ_{j-\lambda} \int_{-u}^{Z_{j-\lambda}} dZ_{j-\lambda-1} \cdots \int_{-u}^{Z_2} dZ_1$$
$$\cdot \int_{-Z_1}^{u} dZ_j \int_{(n-j)/n}^{1-Z_j} dZ_n \int_{(n-j-1)/n}^{Z_n} dZ_{n-1} \cdots \int_{1/n}^{Z_{j+2}} dZ_{j+1}.$$

By the linear transformation

$$x_i = \begin{cases} Z_{j+i} & \text{for } i = 1, 2, \cdots, n - j, \\ 1 - Z_j & \text{for } i = n - j + 1, \\ 1 + Z_{i-n+j-1} & \text{for i} = n - j + 2, n - j + 3, \cdots, n, \end{cases}$$

one obtains

$$K(u, j) = n! \int_{(n-1)/n}^{1} dx_n \cdots \int_{(n-\lambda)/n}^{x_{n-\lambda+2}} dx_{n-\lambda+1} \int_{1-u}^{x_{n-\lambda+1}} dx_{n-\lambda} \cdots \tag{12}$$
$$\int_{1-u}^{x_{n-j+2}} dx_{n-j+1} \int_{(n-j)/n}^{x_{n-j+1}} dx_{n-j} \cdots \int_{2/n}^{x_3} dx_2 \int_{1/n}^{x_2} dx_1.$$

Denote by J_k the result of integration up to and including that with respect to x_k. By repeated integration one finds

$$J_{n-j} = \frac{x_{n-j+1}^{n-j}}{(n - j)!} - \frac{x_{n-j+1}^{n-j-1}}{n(n - j - 1)!}.$$

Hence

$$J_{n-j+1} = \frac{x_{n-j+2}^{n-j+1}}{(n-j+1)!} - \frac{x_{n-j+2}^{n-j}}{n(n-j)!} + W_j(u, j),$$

where for $i = 0, 1, 2, \cdots, j$,

$$W_i(u, j) = \frac{(1-u)^{n-i}}{n(n-i+1)!}(nu - i + 1).$$

Repeated integration gives

$$J_{n-\lambda} = \frac{x_{n-\lambda+1}^{n-\lambda}}{(n-\lambda)!} - \frac{x_{n-\lambda+1}^{n-\lambda-1}}{n(n-\lambda-1)!} + \sum_{i=\lambda+1}^{j} W_i(u, j)\frac{(x_{n-\lambda+1} - 1 + u)^{i-\lambda-1}}{(i-\lambda-1)!}.$$

By properties of the binomial expansion one obtains

$$\frac{x_{n-\lambda+1}^{n-\lambda}}{(n-\lambda)!} - \frac{x_{n-\lambda+1}^{n-\lambda-1}}{n(n-\lambda-1)!} = \frac{-1}{(n-\lambda)!}\sum_{s=0}^{n-\lambda}\binom{n-\lambda}{s}(x_{n-\lambda+1} - 1 + u)^s \cdot (1-u)^{n-\lambda-s-1}[u - (s+\lambda)/n]$$

and therefore

$$(13)\qquad J_{n-\lambda} = \frac{1}{(n-\lambda)!}\sum_{s=j-\lambda}^{n-\lambda}\binom{n-\lambda}{s}(x_{n-\lambda+1} - 1 + u)^s \cdot (1-u)^{n-\lambda-s-1}[(s+\lambda)/n - u].$$

The identity

$$\int_{(n-1)/n}^{1} dx_n \int_{(n-2)/n}^{x_n} dx_{n-1} \cdots \int_{(n-\lambda)/n}^{x_{n-\lambda+2}} (u - 1 + x_{n-\lambda+1})^s \, dx_{n-\lambda+1} \cdots dx_{n-1}\, dx_n$$
$$= \frac{s!}{(s+\lambda)!} n^{-(s+\lambda)}\left[(nu)^{s+\lambda} - \sum_{t=0}^{\lambda-1}\binom{s+\lambda}{t}(nu - 1 - t)^{s+\lambda-t}(1+t)^{t-1}\right]$$

is easily proven by induction on λ. Applying (5) one shows that the right side of this identity is equal to

$$\frac{s!n^{-(s+\lambda)}}{(s+\lambda)!}\sum_{t=\lambda}^{s+\lambda}\binom{s+\lambda}{t}(nu - t - 1)^{s+\lambda-t}(1+t)^{t-1}.$$

Hence it follows from (13) that

$$K(u, j) = n!J_n$$
$$= \frac{1}{n}\sum_{s=j-\lambda}^{n-\lambda}\binom{n}{s+\lambda}(1-u)^{n-s-\lambda-1}(s + \lambda - nu)n^{-s-\lambda}\sum_{t=\lambda}^{s+\lambda}\binom{s+\lambda}{t} \cdot (nu - t - 1)^{s+\lambda-t}(t+1)^{t-1},$$

which is the expression in (11) in the case $nu < j$.

With a few minor changes, the above argument may be also used to prove Theorem 2 for $j = 1$ and $j = n$. For example, in the discussion preceding (12)

one has to define $Z_0 = Z_{n+1} = 0$, in (12) $J_0 = 1$, and in (13) $x_{n+1} = 1$. Since $j = 1$ has $\lambda = 0$, the theorem in this case follows directly from (13).

To complete the proof of Theorem 2, it remains to consider the case of $u \geqq j/n$. Since $D_n^+ \geqq 0$ then, by (3), $U^* \leqq i^*/n$. This implies that for $u > j/n$ we have

$$[U^* < u, i^* = j] = [U^* \leqq j/n, i^* = j] = [i^* = j],$$

hence the first statement of (11) is true.

Proof of Theorem 1. We have

$$\begin{aligned} p_j &= P[i^* = j] = P[U^* \leqq j/n, i^* = j] \\ &= \lim_{u \nearrow j/n} K(u, j) \\ &= \frac{1}{n} \sum_{i=j}^{n} \binom{n}{i} (1 - j/n)^{n-i-1} (i - j) n^{-i} \sum_{t=j-1}^{i} \binom{i}{t} (j - t - 1)^{i-t} (t + 1)^{t-1}, \end{aligned}$$

which, after neglecting zero terms and interchanging summations, becomes for $s = i - t$

$$\begin{aligned} p_j &= n^{-n} \sum_{t=j}^{n} \binom{n}{t} (t + 1)^{t-1} \sum_{s=0}^{n-t} \binom{n - t}{s} (j - t - 1)^s (n - j)^{n-t-s-1} (s + t - j) \\ &= n^{-n} \sum_{t=j}^{n} \binom{n}{t} (t + 1)^{t-1} (n - t - 1)^{n-t-1} \end{aligned}$$

by a direct application of the binomial expansion. Setting $i = n - t - 1$ for $t < n$, one obtains

$$p_j = n^{-n}(n + 1)^{n-1} - n^{-n} \sum_{i=0}^{n-j-1} \frac{1}{i + 1} \binom{n}{i} i^i (n - i)^{n-i-1},$$

the last sum being zero for $j = n$. By Corollary 1 it follows that for all j,

$$p_j = \frac{1}{n} \sum_{i=n-j}^{n-1} \frac{1}{i + 1} \binom{n}{i} \left(\frac{i}{n}\right)^i \left(i - \frac{i}{n}\right)^{n-i-1}.$$

This completes the proof of Theorem 1.

Proof of Theorem 3. With $\lambda = [nu]$ as above, it follows from Theorem 2 that

$$\begin{aligned} H^*(u) &= \sum_{j=1}^{n} K(u, j) \\ &= \sum_{j=1}^{\lambda} p_j + \sum_{j=\lambda+1}^{n} \frac{1}{n} \sum_{i=j}^{n} \binom{n}{i} (1 - u)^{n-i-1} (i - nu) n^{-i} \\ &\quad \cdot \sum_{t=\lambda}^{i} \binom{i}{t} (nu - t - 1)^{i-t} (t + 1)^{t-1}. \end{aligned}$$

Interchanging summations in the last term according to the pattern

$$\sum_{j=\lambda+1}^{n} \sum_{i=j}^{n} \sum_{t=\lambda}^{i} = \sum_{i=\lambda+1}^{n} (i - \lambda) \sum_{t=\lambda}^{i} = \sum_{t=\lambda}^{n} \sum_{i=t}^{n} (i - \lambda)$$

(the second step follows because the index j does not appear in the summand; the last step follows since at $i = \lambda$ the summand is zero), one obtains

$$H^*(u) = \sum_{j=1}^{\lambda} p_j + n^{-n} \sum_{t=\lambda}^{n} \binom{n}{t} (t+1)^{t-1} \tag{14}$$
$$\cdot \sum_{i=t}^{n} (i-\lambda)(i-nu) \binom{n-t}{i-t} (n-nu)^{n-i-1}(nu-t-1)^{i-t}.$$

Using known properties of the binomial expansion, one can show that, whenever $n - t \neq 1$

$$\sum_{s=0}^{n-t} \{(t-\lambda)(t-nu) + s(2t-\lambda-nu) + s^2\} \tag{15}$$
$$\cdot \binom{n-t}{s} (nu-t-1)^s (n-nu)^{n-t-s-1} = -(t-\lambda)(n-t-1)^{n-t-1}.$$

When $n - t = 1$, this sum reduces to

$$\sum_{s=0}^{1} \{(n-1-\lambda)(n-1-nu) + s(2n-2-\lambda-nu) + s^2\}(-1)^s \tag{16}$$
$$= -(n-1-\lambda) - n(1-u).$$

Substituting (15) and (16) into (14), while setting $i - t = s$, one obtains

$$H^*(u) = \sum_{j=1}^{\lambda} p_j - n^{-n} \sum_{t=\lambda}^{n-1} \binom{n}{t} (t-\lambda)(t+1)^{t-1}(n-t-1)^{n-t-1} \tag{17}$$
$$- (1-u) - n^{-n}(n-\lambda)(n+1)^{n-1}.$$

Employing Theorem 1, Corollary 1 with $i = n - t - 1$, and Lemma 2, one concludes from (17)

$$H^*(u) = n^{-n} \sum_{i=n-\lambda}^{n-1} \frac{\lambda - n + i + 1}{i+1} \binom{n}{i} i^i (n-i)^{n-i} - n^{-n} \sum_{i=0}^{n-\lambda-1} \frac{n-i-1-\lambda}{i+1} \cdot$$
$$\binom{n}{i} i^i (n-i)^{n-i-1} - (1-u) - (n-\lambda)p_n$$
$$= n^{-n} \sum_{i=0}^{n-1} \binom{n}{i} i^i (n-i)^{n-i} - 1 + u$$
$$= u.$$

This completes the proof of Theorem 3.

A consequence of Theorem 1 is the following

COROLLARY 2. *For all integers $n > 0, j > 0$,*

$$0 < p_1 < p_2 < \cdots < p_n < 1,$$
$$\lim_{n\to\infty} np_j = \sum_{i=1}^{j} \frac{e^{-i} i^{i-1}}{i!},$$
$$\lim_{n\to\infty} np_{n-j} = e - \sum_{i=0}^{j-1} \frac{e^{-i} i^i}{(i+1)!}.$$

PROOF. The first statement is evident from (10); the second follows from (10) by applying Stirling's formula; and the third follows by applying Stirling's formula to the expression

$$p_{n-j} = \frac{(n+1)^{n-1}}{n^n} - \sum_{i=0}^{j-1} \frac{1}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1},$$

which can be obtained from (10) by Corollary 1.

Thus the statistic D_n^+ places more weight upon the larger observations than on the smaller ones, in the sense that the maximum deviation between F and F_n is more probable to occur at X_{k+1} than at X_k for $k = 1, 2, \cdots, n-1$.

4. The asymptotic distribution of $\alpha_n = i^*/n$. Writing U_n^* instead of U^*, we have according to Theorem 3,

$$P[U_n^* \leqq u] = H_n^*(u) = u, \qquad 0 \leqq u \leqq 1. \tag{18}$$

Since the Glivenko-Cantelli theorem ([3], p. 260) implies that D_n^+ converges in probability to zero, it follows from (3) that

$$\alpha_n - U_n^* \to 0 \quad \text{in probability}. \tag{19}$$

From (18) and (19) one can conclude that α_n is asymptotically uniformly distributed on [0, 1].

The following theorem contains more specific statements on the asymptotic behavior of the distribution of α_n.

THEOREM 4. *For every positive integer n we have*

$$E(\alpha_n) = \frac{1}{2} + \frac{1}{2} n^{-n-1} n! \sum_{i=0}^{n-1} \frac{n^i}{i!}, \tag{20}$$

$$x - \sqrt{n^{-n-1} n! \sum_{i=0}^{n-1} \frac{n^i}{i!}} \leqq \Pr\{\alpha_n < x\} \leqq x \quad \text{for } 0 \leqq x \leqq 1. \tag{21}$$

PROOF OF THEOREM 4. From Theorem 1 we have

$$\begin{aligned} E(\alpha_n) &= n^{-n-1} \sum_{j=1}^{n} \sum_{i=n-j}^{n-1} \frac{j}{i+1} \binom{n}{i} i^i (n-i)^{n-i-1} \\ &= \frac{1}{2}\left(1 - \frac{1}{n}\right) + \frac{1}{2} n^{-n-1} \sum_{i=0}^{n} \binom{n}{i} i^i (n-i)^{n-i} \end{aligned}$$

and this by Lemma 1 yields (20). To obtain the upper bound on $\Pr\{\alpha_n < x\}$ in (21) we note that

$$G_n(x) = \Pr\{\alpha_n < x\} = \sum_{u=1}^{[nx]} p_j$$

and in view of Corollary 2 this must be $<x$ for all $1/n < x \leqq 1$.

To obtain the lower inequality in (21) we need

LEMMA 4. *Let X be a random variable with c.d.f. F, such that $F(0) = 0$, $F(1 + 0) = 1$, $F(x) \leqq x$ for $0 \leqq x \leqq 1$. Then*

$$F(s) \geqq s - \sqrt{2E(X) - 1}. \tag{22}$$

PROOF OF LEMMA 4. We have

$$E(X) = \int_0^1 X \, dF(X) \geqq 1 - \int_0^1 F(X) \, dX$$

$$\geqq 1 - \int_0^{F(s)} X \, dX - \int_{F(s)}^{s} F(s) \, dX - \int_s^1 X \, dX = \tfrac{1}{2}\{1 + [s - F(s)]^2\}$$

and this implies (22). One verifies directly that, for given s and $F(s)$, equality is attained in (22) when $F(t) = t$ for $0 \leqq t \leqq F(s)$, $F(t) = F(s)$ for $F(s) \leqq t \leqq s$, $F(t) = t$ for $s \leqq t \leqq 1$.

According to the upper inequality in (21), Pr $\{\alpha_n < x\}$ fulfills the assumptions of Lemma 4, which together with (20) yields the lower bound of (21).

It may be noted that by an application of Stirling's formula one obtains from (21)

$$0 \leqq x - \Pr\{\alpha_n < x\} = 0(n^{-1/4}), \tag{23}$$

and that (20) together with (3) yields

$$E(D_n^+) = 2^{-1}n^{-n-1}n! \sum_{i=0}^{n-1} \frac{n^i}{i!}. \tag{24}$$

REFERENCES

[1] A. WALD AND J. WOLFOWITZ, "Confidence limits for continuous distribution functions," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 105–118

[2] N. H. ABEL, *Oevres Complètes*, Christiania, C. Groendahl, 1839

[3] M. FRÉCHET, *Recherches théoriques modernes sur la théorie des probabilités*, Gauthier-Villars, Paris, 1937, p. 260.

ON SEVERAL STATISTICS RELATED TO EMPIRICAL DISTRIBUTION FUNCTIONS[1]

BY MEYER DWASS

Northwestern University

1. Introduction. Let $X_1, \cdots, X_n$ be n independent random variables, each with the same continuous c.d.f., $F(x)$. Let $F_n(x)$ be the empirical c.d.f. of the X_i's. We consider the following random variables,

$$U_n = \mu\{F(t): F_n(t) - F(t) > 0\},$$

$$D_n = \sup_{-\infty < t < \infty} (F_n(t) - F(t)),$$

$$V_n = \inf_{-\infty < t < \infty} \{F(t): F_n(t) - F(t) = D_n\},$$

where $\{F(t): \quad\}$ denotes the set of values of $F(t)$, for which t satisfies the condition after the colon. These are sets in the interval (0, 1). In the definition of U_n, $\mu\{\quad\}$ means Lebesgue measure. Obviously, there is no loss of generality in supposing that the X_i's are uniformly distributed over (0, 1) and hence

$$(1) \qquad \begin{cases} U_n = \mu\{t: F_n(t) - t > 0\}, \\ D_n = \sup_{0 \leqq t \leqq 1} (F_n(t) - t), \\ V_n = \inf_{-\infty < t < \infty} \{t: F_n(t) - t = D_n\}. \end{cases}$$

In [5], Kac showed that as $n \to \infty$, U_n has an asymptotic distribution which is uniform over (0, 1). A stronger result was recently obtained by Gnedenko and Mihalevič in [4] in which they showed that *for every* n, U_n is uniformly distributed. Birnbaum and Pyke in a forthcoming paper [2] show that for every n, V_n is also distributed uniformly over (0.1). The methods of [2] and [4] are computational and the purpose of this note is to derive the uniform distribution of U_n and V_n by a short method which employs results of E. S. Andersen and a well-known relationship between the Poisson process and uniformly distributed random variables. In Sec. 3, a generalization of these results is given.

2. Proof of uniform distribution of U_n and V_n. The proof depends on two sets of facts. The first refers to the Poisson process. By this we mean the stochastic process, $X(t)$, with independent and stationary Poisson distributed increments, defined for $t \geqq 0$ and such that $X(0) = 0$. For this process, it is well known that given that $X(1) = n$, a positive integer, then the conditional distribution of the discontinuity (jump) points, $t_1 \leqq t_2 \leqq \cdots \leqq t_n$ of $X(t)$, $0 \leqq t \leqq 1$, is that

Received February 11, 1957; revised June 18, 1957.

[1] Sponsored by Air Force Office of Scientific Research AFOSR-TN-57-784, AD148015 Contract No. AF 49(638)-151.

of the ordered values of n independent, uniform random variables. Another way of saying this, somewhat roughly, is that the conditional distribution of the random function $X(t)$, $0 \leqq t \leqq 1$, given that $X(1) = n$, is that of the empirical c.d.f. of n independent, uniform random variables. For a proof of these facts see p. 400 of [3]. The second set of needed facts is contained in a paper of E. S. Andersen [1], namely:

LEMMA (ANDERSEN). *Let $Y_1, Y_2, \cdots$ be independent and identically distributed random variables. Make the definitions*

$$S_0 = 0 \text{ (a.s.)}, \qquad S_i = \sum_{j=1}^{i} X_j,$$

$$L_r = \textit{smallest } i \textit{ for which } S_i = \max(0, S_1, \cdots, S_r).$$

$$N_r = \textit{number of positive terms in } S_1, \cdots, S_r.$$

Then

$$(2) \qquad P(L_r = m \mid S_{r+1} = 0) = P(N_r = m \mid S_{r+1} = 0) = \frac{1}{r+1},$$

for $m = 0, 1, \cdots, r$ *if and only if*

$$(3) \qquad P(S_i = S_{r+1} = 0) = 0, \qquad (i = 1, 2, \cdots, r).$$

We remark that Andersen's results are much more general, but we state them in a form convenient for our applications.

THEOREM 1. U_n *and* V_n *are each distributed uniformly over* (0, 1).

PROOF. Consider the Poisson process $X(t)$, $0 \leqq t \leqq 1$. Divide the interval (0, 1) into the $r + 1$ parts $(0, 1/(r+1))$, $(1/(r+1), 2/(r+1))$, $\cdots$, $(r/(r+1), 1)$, where $r + 1$ is greater than n and *is a prime number*. (Whenever we state $r \to \infty$ we will understand that $r + 1$ goes through the primes.) The increments of $X(t)$ in these intervals are independent and identically distributed Poisson random variables. We denote these increments by $W_1, W_2, \cdots, W_{r+1}$, respectively, and define $Y_i = W_i - n/(r+1)$, $i = 1, \cdots, r+1$. The Y_i's are independent and identically distributed. We want to show that they satisfy (3) of Andersen's lemma. This is so because $S_i = S_{r+1} = 0$ implies that $(r+1) \cdot X(i/(r+1)) = ni$. This cannot hold since by the primeness of $r + 1$, n must be a factor of $X(i/r+1)$, but since $X(t)$ is non-decreasing this would mean $X(i/(r+1)) = n$, or $r + 1 = i$, a contradiction; thus (3) holds. Under the condition $X(1) = n$, $X(t)$ is distributed like $F_n(t)$, for $s \leqq t \leqq 1$. Hence we can define U_n, V_n for $X(t)$, $0 \leqq t \leqq 1$. We next observe that when $X(1) = n$, then

$$(4) \qquad \left| U_n - \frac{N_r}{r+1} \right| < \frac{A}{r+1}, \qquad \left| V_n - \frac{L_r}{r+1} \right| < \frac{B}{r+1},$$

where A, B are constants which depend on n but not on r. Thus, under the condition $X(1) = n$, both absolute values in (4) converge in probability to zero as $r \to \infty$. Since $N_r/(r+1)$ and $L_r/(r+1)$ are asymptotically uniformly distributed over (0, 1) as $r \to \infty$, this completes the proof.

3. Generalization. A generalization of Theorem 1 can be given which may be of interest. Let

$$X_{11}, \cdots, X_{1n_1}; \cdots; X_{k1}, \cdots, X_{kn_k},$$

be $n = n_1 + \cdots + n_k$ independent random variables each uniformly distributed over (0, 1). Let $F^{(1)}(t), \cdots, F^{(k)}(t)$ be the empirical c.d.f.'s of each of the k sets of variables and define

$$F_\rho(t) = \rho_1 F^{(1)}(t) + \cdots + \rho_k F^{(k)}(t), \qquad 0 \leqq t \leqq 1,$$

where $\rho = (\rho_1, \rho_2, \cdots, \rho_k)$, $\rho_i > 0$, $\rho_1 + \rho_2 + \cdots + \rho_k = 1$. In the special case where $\rho_i = n_i/n$, $i = 1, \cdots, k$, then $F_\rho(t)$ is the empirical c.d.f. of the combined set of n variables. Otherwise $F_\rho(t)$ can only be described as a nondecreasing random step function on (0, 1) such that $F_n(0) = 0$, $F_n(1) = 1$. Nevertheless, random variables U_ρ, D_ρ and V_ρ analogous to U_n, D_n and V_n may be defined for $F_\rho(t)$ exactly as was done in (1) for $F_n(t)$; (replace $F_n(t)$ by $F_\rho(t)$ in (1)). In the following theorem we understand them to be so defined.

THEOREM 2. *U_ρ and V_ρ are each distributed uniformly over* (0, 1).

PROOF. Let $X_1(t), X_2(t), \cdots, X_k(t)$ be k independent Poisson processes and define $X(t) = \rho_1 X_1(t) + \cdots + \rho_k X_k(t)$. Then $X(t)$ is also a process with stationary independent increments. Define now $\rho = (\rho_1, \rho_2, \cdots, \rho_k)$,

$$\begin{cases} \tilde{U}_\rho = \mu\{t: X(t) - X(1)t > 0, 0 \leqq t \leqq 1\}, \\ \tilde{D}_\rho = \sup_{0 \leqq t \leqq 1} (X(t) - X(1)t), \\ \tilde{V}_\rho = \inf_{0 \leqq t \leqq 1} \{t: X(t) - X(1)t = \tilde{D}_\rho\}. \end{cases}$$

We suppose first that

$$(5) \qquad \rho_1 = a_1/a, \cdots, \rho_k = a_k/a,$$

where $a_1, \cdots, a_k$ are positive integers, and $a_1 + \cdots + a_k = a$. If b is a number such that $P(X(1) = b) > 0$, then $\tilde{U}_\rho$, $\tilde{V}_\rho$ are each uniformly distributed over (0, 1) given that $X(1) = b$. The proof of this fact follows exactly the proof of theorem 1. In particular the definition of the ρ_i's by (5) allows a verification of the condition (3) of Andersen's lemma which is exactly analogous to that done in the proof of Theorem 1. Since the ρ_i's as defined by (5) are dense in the set of all possible ρ_i's, it follows by a simple continuity argument that the conditional distribution of $\tilde{U}_\rho$, $\tilde{V}_\rho$ given that $X(1) = b$, is uniform *without* the restriction (5). If $X(1) = \rho_1 X_1(1) + \cdots + \rho_k X_k(1) = b$, this need not uniquely determine the values of the $X_i(1)$. That is, there may be two different sets of positive or zero integers, $x_1, \cdots, x_k$; $y_1, \cdots, y_k$, such that

$$\rho_1 x_1 + \cdots + \rho_k x_k = \rho_1 y_1 + \cdots + \rho_k y_k = b.$$

On the other hand, there is a dense subset of the k-dimensional unit cube where this cannot happen, namely any dense subset, each point of which has rationally

independent coordinates. Thus, in such a dense subset $X(1) = \rho_1 n_1 + \cdots + \rho_k n_k$ if and only if $x_i(1) = n_1, \cdots, x_k(1) = n_k$, for a set of ρ_i's which are dense in the set of all possible ρ_i's. For such ρ_i's the conditional distribution of $\bar{U}_\rho$ and $\bar{V}_\rho$ given that $X_1(1) = n_1, \cdots, X_k(1) = n_k$, is thus uniform. This holds also for the exceptional ρ_i's by a continuity argument. This completes the proof since $F^{(1)}(t), \cdots, F^{(k)}(t)$ are distributed like $X_1(t), \cdots, X_k(t)$ for $0 \leqq t \leqq 1$, under the conditions that $X_1(1) = n_1, \cdots, X_k(t) = n_k$.

4. Concluding remarks. The linear combinations of Theorem 2 are convex ($\rho_1 + \cdots + \rho_k = 1$) and positive ($\rho_i > 0$). The convexity, as well as the strict positivity, is a matter of convenience. The condition of non-negativeness, however, cannot be removed. It is easy to verify directly, for example, that the theorem does not hold for

$$F_\rho(t) = \rho_1 F^{(1)}(t) + \rho_2 F^{(2)}(t),$$

if $\rho_1 > 0$ and $\rho_2 < 0$. The trouble arises because the condition (3) of Andersen's lemma fails to hold.

REFERENCES

[1] E. S. ANDERSEN, "On the fluctuations of sums of random variables," *Math Scand.*, Vol. 1 (1953), pp. 263–285.
[2] Z W. BIRNBAUM AND R PYKE, "On some distributions related to the statistic D_n^+," Technical report no. 23, July 17, 1956, Laboratory of Statistical Research, Dept. of Math., University of Washington, Seattle, Washington.
[3] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York, (1953).
[4] B. V. GNEDENKO AND V. S. MIHALEVIČ, "Two theorems on the behavior of empirical distribution functions," *Doklady Akad Nauk SSSR* (N S.), Vol. 85 (1952), pp. 25–27.
[5] M. KAC, "On deviations between theoretical and empirical distributions," *Proc Nat Acad. Sci., U.S.A.*, Vol. 35 (1949), pp 252–257.

ON THE EFFICIENCY OF ESTIMATES OF TREND IN THE ORNSTEIN UHLENBECK PROCESS

By Charlotte T. Striebel

University of California, Berkeley

1. Summary. The problem is that of estimating the trend of a normal process when the trend function is known up to a finite number of coefficients. That is,

$$y_t = x_t + f(t), \qquad 0 \leqq t \leqq T,$$

where x_t is a normal process with mean zero and covariance function

$$E[X_u, X_v] = C(u, v)$$

and

$$f(t) = k_1\phi_1(t) + \cdots + k_s\phi_s(t).$$

The $\phi_i(t)$ are known functions and the k_i are to be estimated.

The standard procedure in such a case is to derive the estimates by the maximum likelihood method. However, if the covariance function $C(u, v)$ is not completely known, this is usually impossible, and it is essential to find an alternative procedure. The method of least squares has been proposed by Mann [1]. The estimates obtained by this method are independent of $C(u, v)$ and have the additional advantage of being easily computed. Mann and Moranda [2] showed that for the Ornstein Uhlenbeck process the asymptotic efficiency of the least square estimate relative to the maximum likelihood estimate is one, in the special case that the $\phi_i(t)$ are polynomials or trigonometric polynomials. Mann defines the efficiency $\bar{e}(T)$ of an estimate $\bar{f}(t)$

$$\bar{e}(T) = \frac{E\left[\int_0^T [\hat{f}(t) - f(t)]^2\, dt\right]}{E\left[\int_0^T [\bar{f}(t) - f(t)]^2\, dt\right]},$$

where $\hat{f}(t)$ is the maximum likelihood estimate. For the cases that shall be of particular interest—the Ornstein Uhlenbeck process with $\bar{f}(t)$ a linear unbiased estimate—Mann and Moranda [2] have shown that $\bar{e}(T) \leqq 1$.

In the present paper the asymptotic efficiency of the least square estimates will be computed for a wider class of functions $\phi_i(t)$. It will be shown that except for a special case just slightly broader than the one treated by Mann and Moranda, the asymptotic efficiency is actually less than one. Thus except for this special case, the least square estimates could be improved upon. An alternative estimate $\bar{k}_i(\alpha)$ is proposed. It will be shown that for $a \geqq \beta$, where β is the true correlation parameter in the Ornstein Uhlenbeck process, the estimates $\bar{k}_i(\alpha)$ are

Received April 19, 1957; revised October 31, 1957.

asymptotically more efficient than the least square estimates, and in fact as $\alpha \to \beta$ from above the efficiency increases (strictly) to one.

2. Introduction. The least square estimate is obtained by minimizing the expression

$$\int_0^T (y_t - f(t))^2 \, dt$$

and is given by

$$\bar{k}_i = \sum_{j=1}^{s} G^{ij}(T) \int_0^T \phi_i(t) y_t \, dt,$$

where

$$G_{ij}(T) = \int_0^T \phi_i(t)\phi_j(t) \, dt. \tag{1}$$

The maximum likelihood estimates $\hat{k}_i$ minimize

$$\int_0^T \int_0^T [y_u - f(u)][y_v - f(v)] C^{-1}(u, v) \, du \, dv$$

and are given by

$$\hat{k}_i = \sum_{j=1}^{s} \Phi^{ij}(T) \int_0^T \int_0^T \phi_j(u) y_v C^{-1}(u, v) \, du \, dv,$$

where

$$\Phi_{ij}(T) = \int_0^T \int_0^T \phi_i(u)\phi_j(v) C^{-1}(u, v) \, du \, dv. \tag{2}$$

It will be assumed that the $\phi_i(t)$ and $C(u, v)$ are such that these integrals exist. The efficiency of the least square estimates can now be computed

$$\bar{e}(T) = \frac{t[G(T)\Phi^{-1}(T)]}{t[\Psi(T)G^{-1}(T)]},$$

where

$$\Psi(T) = \int_0^T \int_0^T \phi_i(u)\phi_j(v) C(u, v) \, du \, dv. \tag{3}$$

The trace of the matrix is t.

It will further be assumed that there are functions $H_i(T)$ such that the limits

$$\begin{aligned} \lim_{T\to\infty} \frac{G_{ij}(T)}{H_i(T)H_j(T)} &= G_{ij}, \\ \lim_{T\to\infty} \frac{\Phi_{ij}(T)}{H_i(T)H_j(T)} &= \Phi_{ij}, \\ \lim_{T\to\infty} \frac{\Psi_{ij}(T)}{H_i(T)H_j(T)} &= \Psi_{ij} \end{aligned} \tag{4}$$

exist and are positive definite matrices. The asymptotic efficiency then is

$$\bar{e} = \lim_{T\to\infty} \bar{e}(T) = \frac{t(G\Phi^{-1})}{t(\Psi G^{-1})}. \tag{5}$$

Necessary and sufficient conditions that $\bar{e} = 1$ will be found for two classes of G, Φ, Ψ. The first, which includes the cases treated by Mann and Moranda, requires that G, Φ, Ψ be of the form

$$\begin{aligned} G &= \sum_{n=1}^{N} G_n, \\ \Phi &= \sum_{n=1}^{N} c_n G_n, \\ \Psi &= \sum_{n=1}^{N} \frac{1}{c_n} G_n, \end{aligned} \tag{6}$$

where the G_n are positive semi-definite matrices and the c_n are distinct positive real numbers. The second requires that

$$\begin{aligned} \Phi &= BGB^T \\ \Psi &= B^{-1}GB^{-1T} + C, \end{aligned} \tag{7}$$

where B is positive definite, and C is positive semi-definite.

Results will be applied to the case that x_t is an Ornstein Uhlenbeck process and the $\phi_i(t)$ are of the form first

$$\phi_i(t) = \sum_{n=1}^{N} \sum_{r=1}^{\gamma_i} t^r(a_{inr} \sin \omega_n t + b_{inr} \cos \omega_n t)$$

and second

$$\phi_i(t) = e^{a_i t}, \qquad a_i > 0.$$

When the covariance $C(u, v)$ involves some unknown parameters an attempt can be made to estimate them along with the k_i by the maximum likelihood method. However, this frequently leads to equations which cannot be solved. In this case, a natural procedure is to make an estimate $C^*(u, v)$ of $C(u, v)$ by any convenient method and then use the maximum likelihood estimates of the k_i based on the covariance $C^*(u, v)$.

For the Ornstein Uhlenbeck process

$$C(u, v) = \sigma^2 e^{-\beta|u-v|}.$$

Let

$$C^*(u, v) = \begin{cases} \sigma^2 & \text{if } u = v, \\ 0 & \text{otherwise.} \end{cases}$$

This covariance function yields the least square estimates. If the true value β is replaced by α, a family of estimates is obtained by this method.

$$\bar{k}_i(\alpha) = \sum_{j=1}^{s} \Phi_\alpha^{ij}(T) \frac{1}{2}\left[\phi_j(T)y_T + \phi_j(0)y_0 + \frac{1}{\alpha}\int_0^T \phi_j'(t)\,dy_t + \alpha\int_0^T \phi_j(t)y_t\,dt\right], \tag{8}$$

where

$$\Phi_{\alpha ij}(T) = \frac{1}{2}\left[\phi_i(T)\phi_j(T) + \phi_i(0)\phi_j(0) + \frac{1}{\alpha}\int_0^T \phi_i'(t)\phi_j'(t)\,dt + \alpha\int_0^T \phi_i(t)\phi_j(t)\,dt\right]. \tag{9}$$

Clearly

$$\lim_{\alpha\to\infty} \bar{k}_i(\alpha) = \bar{k}_i,$$

$$\bar{k}_i(\beta) = \hat{k}_i.$$

3. Efficiency of estimates for G, Φ, Ψ, of form (6). Assume that G, Φ, Ψ, defined by (1), (2), (3), and (4) are the special form (6). Then from (5)

$$1 - \bar{e} = \frac{t(\Psi G^{-1} - G\Phi^{-1})}{t(\Psi G^{-1})} = \frac{t[G^{-1}\Phi^{-1}(\Phi\Psi - GG)]}{t(\Psi G^{-1})},$$

$$\begin{aligned}(\Phi\Psi - GG) + (\Phi\Psi - GG)^T &= \sum_{n=1}^{N}\sum_{m=1}^{N}\left[\frac{c_n}{c_m}G_nG_m - G_nG_m + \frac{c_m}{c_n}G_nG_m - G_nG_m\right] \\ &= \sum_{n=1}^{N}\sum_{m=1}^{N}\frac{(c_m - c_n)^2}{c_nc_m}G_nG_m.\end{aligned} \tag{10}$$

G_mG_n are positive semi-definite and

$$\frac{(c_m - c_n)^2}{c_nc_m} > 0, \qquad \text{all} \quad m \neq n.$$

Thus, $\Phi\Psi - GG$ is positive semi-definite. In order that $1 - \bar{e} = 0$, it is necessary and sufficient that

$$\Phi\Psi - GG = 0.$$

This is equivalent to requiring

$$t(G_mG_n) = 0, \qquad \text{all } m \neq n.$$

This result will be stated as a theorem.

THEOREM 1. *If G, Φ, Ψ are nonsingular and of the form*

$$G = \sum_{n=1}^{N} G_n ,$$

$$\Phi = \sum_{n=1}^{N} c_n G_n ,$$

$$\Psi = \sum_{n=1}^{N} \frac{1}{c_n} G_n ,$$

where the G_n are positive semi-definite matrices and the c_n are distinct positive real numbers, then

$$\bar{e} = \frac{t(G\Phi^{-1})}{t(\Psi G^{-1})} = 1$$

if and only if

$$t(G_m G_n) = 0, \qquad \text{all } m \neq n.$$

For the Ornstein Uhlenbeck process the theorem can be applied to obtain the special result.

THEOREM 2. *Let*

$$y_t = x_t + f(t),$$

where x_t is an Ornstein Uhlenbeck process with mean zero, and

$$f(t) = k_1\phi_1(t) + \cdots + k_s\phi_s(t).$$

Suppose

$$\phi_i(t) = \sum_{n=1}^{N} \sum_{r=1}^{\gamma_i} t^r (a_{inr} \sin \omega_n t + b_{inr} \cos \omega_n t)$$

are such that

$$\phi_i^*(t) = t^{\gamma_i} \sum_{n=1}^{N} (a_{in\gamma_i} \sin \omega_n t + b_{in\gamma_i} \cos \omega_n t)$$

are linearly independent. Then the asymptotic efficiency of the least square estimates of the k_j is one, if and only if

$$\begin{aligned} \sum a_{in\gamma} a_{im\gamma} &= 0, \\ \sum a_{in\gamma} b_{im\gamma} &= 0, \qquad (11) \\ \sum b_{in\gamma} b_{im\gamma} &= 0, \end{aligned}$$

for all γ and $m \neq n$. The sums extend over all i for which $\gamma_i = \gamma$.

PROOF. Let $H_i(T) = T^{\gamma_i + 1/2}$. The only terms which appear in the limits (4)

will be those of maximum order, that is, those of the $\phi^*(t)$. Denote $a_{in\gamma_i}$ by a_{in} and $b_{in\gamma_i}$ by b_{in}. Then G, Φ, Ψ can be computed and are of the form (6) with

$$G_{nij} = \frac{a_{in}\, a_{jn} + b_{in}\, b_{jn}}{(\gamma_i + \gamma_j + 1)\gamma_i\,!\gamma_j\,!},$$

$$c_n = \frac{\beta^2 + \omega_n^2}{2\beta}.$$

The G_n can easily be shown to be positive semi-definite. Thus by theorem 1, $\bar{e} = 1$ if and only if

$$t(G_m G_n) = 0, \qquad \text{all } m \neq n.$$

$$t(G_m G_n) = \sum_{i=1}^{s}\sum_{j=1}^{s} \frac{a_{in}\, a_{im}\, a_{jn}\, a_{jm} + a_{in}\, b_{im}\, a_{jn}\, b_{jm} + a_{im}\, b_{in}\, a_{jm}\, b_{jn} + b_{in}\, b_{im}\, b_{jn}\, b_{jm}}{(\gamma_i + \gamma_j + 1)^2(\gamma_i\,!)^2(\gamma_j\,!)^2}$$

$$= \sum_{\gamma}\sum_{\delta} \frac{A_{mn\gamma}\, A_{mn\delta} + C_{mn\gamma}\, C_{mn\delta} + C_{nm\gamma}\, C_{nm\delta} + B_{mn\gamma}\, B_{mn\delta}}{(\gamma + \delta + 1)^2(\gamma!)^2(\delta!)^2}.$$

γ and δ are summed over all distinct values of γ_i and

$$A_{mn\gamma} = \sum a_{in} a_{im},$$

$$B_{mn\gamma} = \sum b_{in} b_{im},$$

$$C_{mn\gamma} = \sum a_{in} b_{im}.$$

The summations extend over all values of i for which $\gamma_i = \gamma$ Since

$$\frac{1}{(\gamma + \delta + 1)^2}$$

is a positive definite matrix for γ, δ ranging over distinct integers,

$$t(G_m G_n) = 0, \qquad \text{all } m \neq n$$

if and only if

$$A_{mn\gamma} = B_{mn\gamma} = C_{mn\gamma} = 0$$

for all γ and $m \neq n$.

Thus, unless the special conditions (11) are satisfied, $\bar{e}$ will be strictly less than one. For example,

$$f(t) = k_1 + k_2 \sin t + k_3 \sin 2t$$

can be estimated efficiently by least squares, but

$$f(t) = k_1 + k_2 (\sin t + \sin 2t)$$

cannot.

Grenander and Rosenblatt [3] in Section 7.6 obtain results very similar to those of Theorem 2.

THEOREM 3. *If y_t and $\phi_i(t)$ are as in the hypothesis of theorem 2, then the asymptotic efficiency $\bar{e}(\alpha)$ of the estimate $\bar{k}_i(\alpha)$ (8) is monotone decreasing from 1 at $\alpha = \beta$ to $\bar{e}$ as $\alpha \rightarrow \infty$. If $\bar{e} \neq 1$, then it is strictly decreasing.*

PROOF. First the efficiency of the $\bar{k}_i(\alpha)$ estimates must be computed.

$$E(T, \alpha) = E\left[\int_0^T (\hat{f}_\alpha(t) - f(t))^2 \, dt\right]$$

$$= t[\Sigma_{\bar{k}_i(\alpha)\bar{k}_j(\alpha)}(T) G_{ij}(T)].$$

$$\Sigma_{\bar{k}_i(\alpha)\bar{k}_j(\alpha)} = E[(\bar{k}_i(\alpha) - k_i)(\bar{k}_j(\alpha) - k_j)]$$

$$= \Phi_\alpha^{-1}(T)\left[\frac{(\alpha^2 - \beta^2)}{4} \Psi_\alpha(T) + \frac{\beta}{\alpha} \Phi_\alpha(T)\right] \Phi_\alpha^{-1}(T).$$

$$\Psi_{\alpha ij}(T) = \int_0^T \int_0^T \left(\phi_i(u) + \frac{1}{\alpha} \phi_i'(u)\right)\left(\phi_j(v) + \frac{1}{\alpha} \phi_j'(v)\right) e^{-\beta|u-v|} \, du \, dv,$$

and $\Phi_\alpha(T)$ is defined by (9).

For $\phi_i(t)$ as in theorem 2 and

$$H_i(T) = T^{\gamma_i + 1/2},$$

$$\Phi_{\alpha ij} = \lim_{T\to\infty} \frac{\Phi_{\alpha ij}(T)}{H_i(T)H_j(T)} = \frac{1}{2\alpha} [\alpha^2 G_{ij} + \zeta_{ij}],$$

and

$$\Phi = \frac{1}{2\beta} (\beta^2 G + \zeta),$$

where

$$\zeta_{ij} = \lim_{T\to\infty} \frac{\int_0^T \phi_i'(t)\phi_j'(t) \, dt}{H_i(T)H_j(T)}.$$

$$\Psi_{\alpha ij} = \lim_{T\to\infty} \frac{\Psi_{\alpha ij}(T)}{H_i(T)H_j(T)} = \frac{(\alpha^2 - \beta^2)}{\alpha^2} \Psi_{ij} + \frac{2\beta}{\alpha^2} G_{ij}.$$

Thus,

$$\Phi_\alpha = \frac{1}{2\alpha} [(\alpha^2 - \beta^2)G + 2\beta\Phi].$$

Let

$$A = (\alpha^2 - \beta^2)G + 2\beta\Phi.$$

Then

$$E(\alpha) = \lim_{T\to\infty} E(T, \alpha)$$

$$= t\{A^{-1}A^{-1}[(\alpha^2 - \beta^2)^2\Psi + 4\beta(\alpha^2 - \beta^2)G + 4\beta^2\Phi]\},$$

and

$$\frac{\partial E(\alpha)}{\partial \alpha} = 8\alpha\beta(\alpha^2 - \beta^2)t[A^{-1}A^{-1}GA^{-1}\Phi(G^{-1}\Psi - \Phi^{-1}G)].$$

It follows from (10) that

$$G^{-1}\Psi - \Phi^{-1}G$$

is positive semi-definite. Thus, for $\alpha > \beta$, the derivative of $E(\alpha)$ is nonnegative and $E(\alpha)$ is monotone increasing. If $\bar{e} \neq 1$, then

$$t(G^{-1}\Psi - \Phi^{-1}G) > 0,$$

and at least one characteristic root must be nonzero Since $A^{-1}A^{-1}GA^{-1}\Phi$ is positive definite $\partial E(\alpha)/\partial\alpha$ will then be positive, and $E(\alpha)$ is strictly increasing.

$$\bar{e}(\alpha) = \lim_{T\to\infty} \frac{E\left[\int_0^T (\hat{f}(t) - f(t))^2\, dt\right]}{E\left[\int_0^T (\hat{f}_\alpha(t) - f(t))^2\, dt\right]}.$$

$$= \frac{t(\Phi^{-1}G)}{E(\alpha)}.$$

4. Efficiency of estimates for exponential $\phi_i(t)$. Assume G, Φ, Ψ are of the form (7). Then

$$1 - \bar{e} = \frac{t(\Psi G^{-1} - \Phi^{-1}G)}{t(\Psi G^{-1})} = \frac{t(CG^{-1})}{t(\Psi G^{-1})},$$

and $\bar{e} = 1$ if and only if $C = 0$.

Let

$$\phi_i(t) = e^{a_i t}$$

and

$$H_i(\alpha) = e^{a_i T},$$

where the a_i are positive and distinct. Then

$$G_{ij} = \frac{1}{a_i + a_j},$$

$$\Phi_{ij} = \frac{(\beta + a_i)(\beta + a_j)}{a_i + a_j},$$

$$\Psi_{ij} = \frac{2\beta}{(\beta + a_i)(\beta + a_j)(a_i + a_j)} + \frac{1}{(\beta + a_i)(\beta + a_j)}.$$

Thus,

$$B_{ij} = \frac{(\beta + a_i)\delta_{ij}}{\sqrt{2\beta}},$$

$$C = \frac{1}{2\beta} B^{-1} B \neq 0,$$

and hence

$$\bar{e} < 1.$$

Since the least square estimates are not asymptotically efficient, it is of interest to compute the efficiency of the $\bar{k}_i(\alpha)$ estimates. In this case

$$\Phi_{\alpha ij} = \frac{(\alpha + a_i)(\alpha + a_j)}{2\alpha(a_i + a_j)} = \frac{1}{2\alpha} AGA,$$

where

$$A_{ij} = (\alpha + a_i)\delta_{ij}$$

$$\Psi_\alpha = \frac{1}{\alpha^2} A\Psi A$$

$$E(\alpha) = t\{G^{-1}A^{-1}GA^{-1}G^{-1}[(\alpha^2 - \beta^2)\Psi + 2\beta G]\}$$

$$\frac{\partial E(\alpha)}{\partial \alpha} = 2(\alpha - \beta)t[A^{-1}G^{-1}D],$$

where

$$D_{ij} = \frac{(\alpha - \beta)a_i + (\alpha + \beta)a_j + 2a_i a_j}{(a_i + a_j)(\alpha + a_i)^2(\beta + a_i)(\alpha + a_j)}.$$

For $\alpha > \beta$ this matrix is positive definite. Thus, $\partial E(\alpha)/\partial\alpha$ is positive, and $E(\alpha)$ is strictly increasing. Thus, for $\alpha > \beta$, the $\bar{k}_i(\alpha)$ estimates are more efficient than the least square estimates.

5. Acknowledements. I am indebted to Professor H. B. Mann for suggesting the problem and to Professor Lucien LeCam for many valuable suggestions.

REFERENCES

[1] H. B. MANN, "A theory of estimation for the fundamental random process and the Ornstein Uhlenbeck process," *Sankhyā*, Vol. 13 (1954), p. 325.

[2] H. B. MANN AND P. B. MORANDA, "On the efficiency of the least square estimates of parameters in the Ornstein Uhlenbeck process," *Sankhyā*, Vol. 13 (1954), p. 351.

[3] U. GRENANDER AND M. ROSENBLATT, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1957.

UNBIASED ESTIMATION OF CERTAIN CORRELATION COEFFICIENTS[1]

By Ingram Olkin[2] and John W. Pratt[3]

University of Chicago

1. Summary and introduction. This paper deals with the unbiased estimation of the correlation of two variates having a bivariate normal distribution (Sec. 2), and of the intraclass correlation, i.e., the common correlation coefficient of a p-variate normal distribution with equal variances and equal covariances (Sec. 3).

In both cases, the estimator has the following properties. It is a function of a complete sufficient statistic and is therefore the unique (except for sets of probability zero) minimum variance unbiased estimator. Its range is the region of possible values of the estimated quantity It is a strictly increasing function of the usual estimator differing from it only by terms of order $1/n$ and consequently having the same asymptotic distribution.

Since the unbiased estimators are cumbersome in form in that they are expressed as series or integrals, tables are included giving the unbiased estimators as functions of the usual estimators.

In Sec. 4 we give an unbiased estimator of the squared multiple correlation. It has the properties mentioned in the second paragraph except that it may be negative, which the squared multiple correlation cannot.

In each case the estimator is obtained by inverting a Laplace transform

We are grateful to W. H. Kruskal and L. J. Savage for very helpful comments and suggestions, and to R. R. Blough for his able computations.

2. Correlation coefficient. Let $(x_1, y_1), \cdots, (x_N, y_N)$ be independently distributed, each bivariate normal with means μ_1, μ_2, variances σ_1^2, σ_2^2 and correlation ρ. The problem is to estimate ρ unbiasedly in the cases (i) μ_1, μ_2 known, $\sigma_1^2, \sigma_2^2, \rho$ unknown, and (ii) all parameters unknown.

Sufficiency and invariance suggest that we confine ourselves to odd functions of r, where r is the usual sample correlation coefficient in either case, namely,

$$r = \frac{\sum (x_i - \hat{\mu}_1)(y_i - \hat{\mu}_2)}{\sqrt{\sum (x_i - \hat{\mu}_1)^2 \sum (y_i - \hat{\mu}_2)^2}},$$

where $(\hat{\mu}_1, \hat{\mu}_2)$ equals (μ_1, μ_2) in (i) and $(\bar{x}, \bar{y})$ in (ii).

2.1. *Derivation of the unbiased estimator.* The density of r is

$$(2.1) \qquad p(r) = \frac{2^{n-2}}{\pi\Gamma(n-1)} (1-\rho^2)^{n/2}(1-r^2)^{(n-3)/2} \sum_{k=0}^{\infty} \Gamma^2\left(\frac{n+k}{2}\right) \frac{(2\rho r)^k}{k!},$$

Received August 20, 1956; revised August 15, 1957

[1] This research was sponsored by the Office of Ordnance Research, U. S Army, and the Office of Naval Research.

[2] Work done while on leave from Michigan State University

[3] Now at Harvard University.

where the degrees of freedom are $n = N$ and $N - 1$ in cases (i) and (ii). (We assume $n \geqq 2$, the case $n = 1$ being degenerate.) The condition $E[G(r)] = \rho$, i.e., $G(r)$ is unbiased, is equivalent to

$$\frac{2^{n-2}}{\pi\Gamma(n-1)} \sum_{k=0}^{\infty} \Gamma^2\left(\frac{n+k}{2}\right) \frac{(2\rho)^k}{k!} \int_{-1}^{1} G(r)(1-r^2)^{(n-3)/2} r^k \, dr = (1-\rho^2)^{-n/2}\rho$$

$$= \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{n}{2}+j\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{\rho^{2j+1}}{j!}.$$

Comparing coefficients of powers of ρ, we find that $G(r)$ is indeed an odd function, and that

$$\int_0^1 G(r)(1-r^2)^{(n-3)/2} r^{2j+1} \, dr = \frac{\pi\Gamma(n-1)\Gamma(2j+2)}{2^{n+2j}\Gamma^2\left(\frac{n+2j+1}{2}\right)} \frac{\Gamma\left(\frac{n+2j}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma(j+1)}.$$

Using the identity (e.g., [3, 12.4.4])

$$\sqrt{\pi}\Gamma(2p) = 2^{2p-1}\Gamma(p)\Gamma(p+1/2),$$

and making the substitution $r = \exp(-\frac{1}{2}y)$, we obtain

$$\int_0^\infty G(e^{-\frac{1}{2}y})(1-e^{-y})^{(n-3)/2} e^{-y} e^{-jy} \, dy = \Gamma\left(\frac{n-1}{2}\right) \frac{\Gamma\left(\frac{3}{2}+j\right)\Gamma\left(\frac{n}{2}+j\right)}{\Gamma^2\left(\frac{n+1}{2}+j\right)}.$$

As a function of j, for $n \geqq 2$, the right-hand side is the unilateral Laplace transform of

$$e^{-\frac{3}{2}y}(1-e^{-y})^{((n-1)/2)-1} F\left(\frac{1}{2}, \frac{1}{2}; \frac{n-1}{2}; 1-e^{-y}\right)$$

[1, p. 262 (7)], where F is the hypergeometric function

(2.2)
$$F(\alpha, \beta; \gamma; x) = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+k)\Gamma(\beta+k)\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma+k)} \frac{x^k}{k!}.$$

Therefore

(2.3)
$$G(r) = rF(\tfrac{1}{2}, \tfrac{1}{2}; (n-1)/2; 1-r^2).$$

Some alternative representations of $G(r)$ are

(2.4)
$$G(r) = r \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \int_0^1 \frac{t^{-1/2}(1-t)^{((n-2)/2)-1}}{[1-t(1-r^2)]^{1/2}} \, dt$$

and

$$(2.5)\qquad G(r) = r\,\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}\int_0^\infty \frac{t^{-1/2}(1+t)^{-(n-2)/2}}{(1+tr^2)^{1/2}}\,dt$$

[2, 2.12 (1) and (5)].

2.2 *Properties of the unbiased estimator.* $G(r)$ is an odd function of r by (2.3), and is strictly increasing since, in (2.5), $r(1 + tr^2)^{-1/2}$ is strictly increasing in r for each value of t, $0 < t < \infty$. For $\rho = \pm 1$, $G(r) = r = \pm 1$ with probability 1, and consequently, $-1 \leqq G(r) \leqq 1$, which is the range of ρ.

As remarked before, $G(r)$ is the unique minimum variance unbiased estimator of ρ.

To obtain the asymptotic distribution of $G(r)$, we note that, by (2.2),

$$F(\alpha, \beta; \gamma; x) = 1 + x\,O(1/\gamma)$$

as $\gamma \to \infty$ (uniformly in x for x in any bounded set), so that $G(r) = r + O_p(1/n)$. Therefore $\sqrt{n}[G(r) - \rho]$ has the same asymptotic distribution as $\sqrt{n}[r - \rho]$, which is $N(O, (1 - \rho^2)^2)$, [3, p. 366].

In order to facilitate the use of the unbiased estimator $G(r)$, Table 1 gives $G(r)$ and (for easier interpolation) $G(r)/r$ for $r = 0(.1)\,1$ and $n = 2(2)\,30$. The computation was carried out by means of the recursive relation

$$xF(\tfrac{1}{2}, \tfrac{1}{2}; \gamma + 1; x)$$
$$= \left[1 - \frac{1}{(2\gamma - 1)^2}\right][(2x - 1)F(\tfrac{1}{2}, \tfrac{1}{2}; \gamma; x) + (1 - x)F(\tfrac{1}{2}, \tfrac{1}{2}; \gamma - 1; x)],$$

[2, 2.8 (30)], together with the initial conditions

$$F(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{1}{2}; x) = 1/\sqrt{1 - x},$$
$$F(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{3}{2}; x) = \text{arc sin}\,\sqrt{x}/\sqrt{x},$$

[2, 2.8 (4) and (13)].

Approximations for $G(r)$ can be obtained from the expansion (2.2), which gives

$$(2.6)\qquad \frac{G(r)}{r} = 1 + \frac{1 - r^2}{2(n-1)} + \frac{9(1 - r^2)^2}{8(n^2 - 1)} + O(n^{-3}).$$

(2.6) gives $G(r)/r$ within .01 for $n \geqq 14$ or .001 for $n \geqq 36$ if two terms are included, and within .01 for $n \geqq 10$ or 001 for $n \geqq 18$ if three terms are included. The neglected terms in the first line of (2.6) are all positive and decreasing in r^2 and n. Therefore, if $G(r)$ is estimated by cutting off this series, the estimate will be too small, by a percentage which decreases as r^2 and n increase.

The k that minimizes the maximum over r of the absolute difference between (2.6) and $1 + (1 - r^2)/2(n - k)$ is, for large n, $(-7 + 9\sqrt{2})/2 = 2.87$. This suggests the approximation

$$(2.7)\qquad \frac{G(r)}{r} = 1 + \frac{1 - r^2}{2(n-3)}.$$

This is accurate within .01 for $n \geqq 8$, and within .001 for $n \geqq 18$.

TABLE 1

Ordinary bivariate correlation coefficient, n degrees of freedom

$$G(r) = rF\left(1/2, 1/2; \frac{n-1}{2}; 1 - r^2\right), \qquad r = \frac{s_{12}}{\sqrt{s_{11}\, s_{22}}}$$

1a. *Table of $G(r)$*

n	r = 0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
2	0	1	1	1	1	1	1	1	1	1	1
4	0	.148	.280	.398	.506	.605	.695	.780	.858	.931	1
6	0	.117	.232	.343	.450	.552	.650	.744	.833	.918	1
8	0	.110	.220	.327	.432	.534	.633	.730	.823	.913	1
10	0	.107	.214	.319	.423	.525	.625	.722	.817	.910	1
12	0	.106	.211	.315	.418	.520	.620	.718	.814	.908	1
14	0	.105	.209	.312	.415	.516	.616	.715	.812	.907	1
16	0	.104	.207	.311	.413	.514	.614	.713	.810	.906	1
18	0	.103	.206	.309	.411	.512	.612	.711	.809	.905	1
20	0	.103	.206	.308	.410	.511	.611	.710	.808	.905	1
22	0	.103	.205	.307	.409	.510	.610	.709	.807	.904	1
24	0	.102	.205	.307	.408	.509	.609	.708	.806	.904	1
26	0	.102	.204	.306	.407	.508	.608	.707	.806	.903	1
28	0	.102	.204	.305	.407	.507	.607	.707	.805	.903	1
30	0	.102	.204	.305	.406	.507	.607	.706	.805	.903	1
∞	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1

1b. *Table of $G(r)/r$*

2	∞	10.000	5.000	3.333	2.500	2.000	1.667	1.429	1.250	1.111	1
4	1.571	1.478	1.398	1.327	1.265	1.209	1.159	1.114	1.073	1.035	1
6	1.178	1.173	1.161	1.144	1.125	1.105	1.083	1.062	1.041	1.020	1
8	1.104	1.103	1.098	1.090	1.080	1.068	1.056	1.042	1.028	1.014	1
10	1.074	1.073	1.070	1.065	1.058	1.050	1.042	1.032	1.022	1.011	1
12	1.057	1.056	1.054	1.050	1.046	1.040	1.033	1.026	1.018	1.009	1
14	1.046	1.046	1.044	1.041	1.038	1.033	1.027	1.021	1.015	1.008	1
16	1.039	1.039	1.037	1.035	1.032	1.028	1.023	1.018	1.013	1.006	1
18	1.034	1.033	1.032	1.030	1.028	1.024	1.020	1.016	1.011	1.006	1
20	1.030	1.029	1.028	1.027	1.024	1.022	1.018	1.014	1.010	1.005	1
22	1.027	1.026	1.025	1.024	1.022	1.019	1.016	1.013	1.009	1.005	1
24	1.024	1.024	1.023	1.022	1.020	1.018	1.015	1.012	1.008	1.004	1
26	1.022	1.022	1.021	1.020	1.018	1.016	1.014	1.011	1.007	1.004	1
28	1.020	1.020	1.019	1.018	1.017	1.015	1.012	1.010	1.007	1.004	1
30	1.019	1.018	1.018	1.017	1.015	1.014	1.012	1.009	1.006	1.003	1
∞	1	1	1	1	1	1	1	1	1	1	1

By (2.2) and (2.3), $G(r)/r$ is larger than 1 and decreasing in r^2 and n, as Table 1b suggests.

2.3 *Partial correlation coefficient.* We observe that an unbiased estimator of the partial correlation coefficient can be immediately obtained from the preceding

section. More precisely, suppose the columns of $X:p \times N$ are independently distributed each as p-variate normal with mean vector μ and covariance matrix Σ. We wish to give an unbiased estimator of the partial correlation coefficient $\rho_{12 \cdot (q \cdots p)}$. The usual estimator, $r_{12 \cdot (q \cdots p)}$, has the density (2.1) with $n = N - (p - q)$ if μ is known, and $n = N - 1 - (p - q)$ if μ is unknown. Therefore $G(r_{12 \cdot (q \cdots p)})$ (with appropriate n) is the unique minimum variance unbiased estimator of $\rho_{12 \cdot (q \cdots p)}$ and possesses the other properties of $G(r)$.

3. Intraclass correlation coefficient. Let the columns of X: $p \times N$ be independently distributed, each as $N(\mu, \Sigma^*)$, i.e., p-variate normal with mean vector μ and covariance matrix Σ^*. Suppose Σ^* is of the form $\sigma^2[(1 - \rho) I + \rho ee']$, where $e' = (1, \cdots, 1)$, i.e., $\sigma^*_{ii} = \sigma^2$, $\sigma^*_{ij} = \rho\sigma^2$ $(i \neq j)$, with ρ and σ^2 unknown. The problem is to estimate ρ unbiasedly.

We note that ρ is just the slope of the regression line of x_2 on x_1, and is therefore estimated unbiasedly by

$$\hat{\rho} = \frac{\sum_{\alpha=1}^{N} (x_{1\alpha} - x_{1\cdot})(x_{2\alpha} - x_{2\cdot})}{\sum_{\alpha=1}^{N} (x_{1\alpha} - x_{1\cdot})^2},$$

where a dot indicates an average over the omitted subscript. We will see presently that there is a complete sufficient statistic (u, v). $\hat{\rho}$ is not a function of (u, v), nor is it confined to the range of ρ, namely, $-1/(p - 1)$ to 1. However, by the Blackwell-Rao theorem, $E(\hat{\rho} \mid u, v)$ is the unique minimum variance unbiased estimator of ρ. Since $E(\hat{\rho} \mid u, v)$ is difficult to obtain, we shall use the joint distribution of u and v to obtain an unbiased estimator $h(u, v)$ of ρ, which, by completeness, must equal $E(\hat{\rho} \mid u, v)$.

As in the previous section, sufficiency and invariance suggest that we confine ourselves to functions of the conventional estimator r' of ρ. However, it is easier to deal with the density of (u, v), and it will turn out that the unbiased estimator $h(u, v)$ is a function $H(r')$ of r' alone.

3.1 *Reduction to canonical form.* Let $\Delta:p \times p$ be an orthogonal matrix with first row $p^{-1/2}e'$, and let $Y = \Delta X$. Then the columns of Y are independently distributed, each as $N(\Delta\mu, \Delta\Sigma^*\Delta')$. Now

$$\Delta\Sigma^*\Delta' = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 I \end{pmatrix},$$

where $\sigma_1^2 = \sigma^2 [1 + (p - 1)\rho]$, $\sigma_2^2 = \sigma^2 (1 - \rho)$. Because of the particular diagonal form of the covariance matrix, the $y_{i\alpha}(i = 1, \cdots, p; \alpha = 1, \cdots, N)$ are independent, and if we let $\eta = \Delta\mu = Ey$, then $y_{1\alpha}$ is $N(\eta_1, \sigma_1^2)$, $(\alpha = 1, \cdots, N)$ and $y_{i\alpha}$ is $N(\eta_i, \sigma_2^2)$, $(i = 2, \cdots, p; \alpha = 1, \cdots, N)$. We can therefore obtain two sums of squares, u and v, sufficient for σ_1^2 and σ_2^2 and distributed independently as $\sigma_1^2 \chi_a^2$ and $\sigma_2^2 \chi_b^2$ where the degrees of freedom a and b depend on our knowledge

of μ. To write u and v conveniently, we first observe that

$$\sum_{i=1}^{p}\sum_{\alpha=1}^{N} y_{i\alpha}^2 = \operatorname{tr} YY' = \operatorname{tr} XX' = \sum_{i=1}^{p}\sum_{\alpha=1}^{N} x_{i\alpha}^2,$$

$$\sum_{\alpha=1}^{N} y_{1\alpha}^2 = p^{-1}e'XX'e = p\sum_{\alpha=1}^{N} x_{\cdot\alpha}^2,$$

$$\sum_{i=1}^{p} y_{i\cdot}^2 = N^{-2}e'Y'Ye = N^{-2}e'X'Xe = \sum_{i=1}^{p} x_{i\cdot}^2,$$

$$y_{1\cdot} = N^{-1}p^{-1/2}e'Xe = p^{1/2}x_{\cdot\cdot}\,.$$

Precisely, we consider the following three cases:

(i) $\mu = 0$ and hence $\eta = \Delta\mu = 0$. Let $a = N$, $b = (p-1)N$,

$$u = \sum_{\alpha=1}^{N} y_{1\alpha}^2 = p\sum_{\alpha=1}^{N} x_{\cdot\alpha}^2,$$

$$v = \sum_{i=2}^{p}\sum_{\alpha=1}^{N} y_{i\alpha}^2 = \sum_{\alpha=1}^{N}\sum_{i=1}^{p} (x_{i\alpha} - x_{\cdot\alpha})^2.$$

(ii) μ completely unknown and hence $\eta = \Delta\mu$ is also completely unknown. Let $a = N-1$, $b = (p-1)(N-1)$,

$$u = \sum_{\alpha=1}^{N} (y_{1\alpha} - y_{1\cdot})^2 = p\sum_{\alpha=1}^{N} (x_{\cdot\alpha} - x_{\cdot\cdot})^2,$$

$$v = \sum_{i=2}^{p}\sum_{\alpha=1}^{N} (y_{i\alpha} - y_{i\cdot})^2 = \sum_{\alpha=1}^{N}\sum_{i=1}^{p} (x_{i\alpha} - x_{i\cdot} - x_{\cdot\alpha} + x_{\cdot\cdot})^2.$$

(iii) $\mu = \omega e$, where ω is an unknown scalar, and hence $\eta = \omega\Delta e = \omega\sqrt{p}(1, 0, \cdots, 0)'$. Let $a = N-1$, $b = (p-1)N$,

$$u = \sum_{\alpha=1}^{N} (y_{1\alpha} - y_{1\cdot})^2 = p\sum_{\alpha=1}^{N} (x_{i\alpha} - x_{\cdot\cdot})^2,$$

$$v = \sum_{i=2}^{p}\sum_{\alpha=1}^{N} y_{i\alpha}^2 = \sum_{i=1}^{p}\sum_{\alpha=1}^{N} (x_{i\alpha} - x_{\cdot\alpha})^2.$$

In each case u/σ_1^2, and v/σ_2^2 are independently distributed as χ_a^2 and χ_b^2, and it is easily shown that (u, v) is a complete sufficient statistic for (σ_1^2, σ_2^2). The three cases can thus be treated simultaneously.

3.2 *Derivation of the unbiased estimator.* The condition that $h(u, v)$ be unbiased is

$$\int_0^\infty\int_0^\infty h(u, v)u^{a/2-1}v^{b/2-1}e^{-\theta u-\phi v}\,du\,dv = \Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{b}{2}\right)\frac{\phi-\theta}{\phi+(p-1)\theta}\cdot\frac{1}{\theta^{a/2}\phi^{b/2}}, \tag{3.1}$$

where $\theta = 1/(2\sigma_1^2)$, $\phi = 1/(2\sigma_2^2)$. The right-hand side is the bivariate Laplace transform of

$$(3.2)\quad \begin{aligned}&\left(\frac{b}{2}-1\right)\int_0^L [u-(p-1)y]^{\frac{a}{2}-1}(v-y)^{\frac{b}{2}-2}\,dy\\ &\qquad -\left(\frac{a}{2}-1\right)\int_0^L [u-(p-1)t]^{\frac{a}{2}-2}(v-t)^{\frac{b}{2}-1}\,dt,\end{aligned}$$

where $L = \min\,[u/(p-1), v]$, [4, p. 36 (Satz 12), p. 208 (9), p. 236 (87)]. Integrating the first term of (3.2) by parts and letting $z = u/[(p-1)v]$, we obtain

$$(3.3)\quad \begin{aligned}h(u, v) = h^*(z) &= 1 - \left(\frac{a}{2}-1\right)\frac{p}{p-1}\int_0^1 (1-zw)^{\frac{b}{2}-1}(1-w)^{\frac{a}{2}-2}\,dw\\ &= 1 - \frac{p}{p-1}F\left(1, 1-\frac{b}{2}; \frac{a}{2}, z\right) \qquad \text{for } 0 \leqq z \leqq 1,\end{aligned}$$

$$(3.4)\quad \begin{aligned}h(u, v) = h^*(z) &= 1 - \left(\frac{a}{2}-1\right)\frac{p}{p-1}\frac{1}{z}\int_0^1 \left(1-\frac{1}{z}w\right)^{\frac{a}{2}-2}(1-w)^{\frac{b}{2}-1}\,dw\\ &= 1 - \frac{2}{b}\left(\frac{a}{2}-1\right)\frac{p}{p-1}\frac{1}{z}F\left(1, 2-\frac{a}{2}; \frac{b}{2}+1; \frac{1}{z}\right)\\ &\qquad \text{for } z \geqq 1,\end{aligned}$$

[2, 2.12 (1)]. Integrating the second term of (3.2) by parts we obtain the following alternative to (3.4):

$$(3.5)\quad \begin{aligned}h^*(z) &= \left(\frac{b}{2}-1\right)\frac{p}{p-1}\int_0^1 \left(1-\frac{1}{z}w\right)^{\frac{a}{2}-1}(1-w)^{\frac{b}{2}-2}\,dw - \frac{1}{p-1}\\ &= \frac{p}{p-1}F\left(1, 1-\frac{a}{2}; \frac{b}{2}; \frac{1}{z}\right) - \frac{1}{p-1} \qquad \text{for } z \geqq 1,\end{aligned}$$

[2, 2.12 (1)].

The conventional estimate of ρ is (e.g., [5] and [6]),

$$(3.6)\quad \begin{aligned}r' &= \frac{p}{p-1}\frac{\sum_{\alpha}\sum_{i\neq j}(x_{i\alpha}-\hat{\mu}_i)(x_{j\alpha}-\hat{\mu}_j)}{\sum_{\alpha}\sum_{i}(x_{i\alpha}-\hat{\mu}_i)^2}\\ &= \frac{1}{p-1}\left(\frac{pu}{u+v}-1\right) = \frac{pz}{1+(p-1)z} - \frac{1}{p-1},\end{aligned}$$

where $\hat{\mu}$ is the appropriate estimate of μ in (i), (ii), or (iii). Now

$$(3.7)\quad z = \frac{(p-1)\,r'+1}{(p-1)^2(1-r')},$$

which is a strictly increasing function of r'. Thus $h^*(z)$ is a function $H(r')$ of r'.

TABLE 2

Intraclass correlation coefficient, bivariate case, n degrees of freedom

$$H(r') = r'F(1/2, 1; n/2; 1 - r'^2)$$

2a. *Table of* $H(r')$

n	r': 0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
2	0	1	1	1	1	1	1	1	1	1	1
4	0	.182	.333	.462	.571	.667	.750	.824	.889	.947	1
6	0	.132	.259	.379	.490	.593	.688	.775	.856	.931	1
8	0	.120	.237	.351	.459	.563	.661	.753	.841	.923	1
10	0	.114	.227	.337	.444	.547	.647	.741	.832	.918	1
12	0	.111	.221	.329	.435	.538	.638	.734	.826	.915	1
14	0	.109	.217	.324	.429	.532	.631	.728	.822	.913	1
16	0	.108	.215	.321	.425	.527	.627	.724	.819	.911	1
18	0	.107	.213	.318	.422	.524	.624	.722	.817	.910	1
20	0	.106	.211	.316	.419	.521	.621	.719	.815	.909	1
22	0	.105	.210	.314	.417	.519	.619	.717	.814	.908	1
24	0	.105	.209	.313	.416	.517	.617	.716	.813	.907	1
26	0	.104	.208	.312	.414	.516	.616	.715	.812	.907	1
28	0	.104	.208	.311	.413	.515	.615	.713	.811	.906	1
30	0	.104	.207	.310	.412	.513	.614	.713	.810	.906	1
∞	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1

2b. *Table of* $H(r')/r'$

2	∞	10.000	5.000	3.333	2.500	2.000	1.667	1.429	1.250	1.111	1
4	2.000	1.818	1.667	1.538	1.429	1.333	1.250	1.176	1.111	1.053	1
6	1.333	1.322	1.296	1.262	1.224	1.185	1.146	1.107	1.070	1.034	1
8	1.200	1.196	1.185	1.169	1.149	1.126	1.102	1.076	1.051	1.025	1
10	1.143	1.141	1.134	1.124	1.110	1.095	1.078	1.059	1.040	1.020	1
12	1.111	1.110	1.105	1.098	1.088	1.076	1.062	1.048	1.033	1.017	1
14	1.091	1.090	1.086	1.080	1.073	1.063	1.052	1.041	1.028	1.014	1
16	1.077	1.076	1.073	1.068	1.062	1.054	1.045	1.035	1.024	1.012	1
[illegible]	1.067	1.066	1.063	1.059	1.054	1.047	1.040	1.031	1.021	1.011	1
[illegible]	1.059	1.058	1.056	1.053	1.048	1.042	1.035	1.027	1.019	1.010	1
[illegible]	1.053	1.052	1.050	1.047	1.043	1.038	1.032	1.025	1.017	1.009	1
4	1.048	1.047	1.045	1.043	1.039	1.034	1.029	1.023	1.016	1.008	1
26	1.043	1.043	1.042	1.039	1.036	1.032	1.027	1.021	1.014	1.007	1
28	1.040	1.040	1.038	1.036	1.033	1.029	1.024	1.019	1.013	1.007	1
30	1.037	1.037	1.035	1.033	1.031	1.027	1.023	1.018	1.012	1.006	1
∞	1	1	1	1	1	1	1	1	1	1	1

3.3. *Properties of the unbiased estimator.* For $\rho = 1$, $z = \infty$, $r' = 1$ with probability 1, and $h^*(\infty) = H(1) = 1$; for $\rho = -1/(p-1)$, $z = 0$, $r' = -1/(p-1)$ with probability 1, and $h^*(0) = H(-1/(p-1)) = -1/(p-1)$. Thus in the two cases when Σ^* is singular, $h^*(z) = H(r') = \rho$ with probability 1. Furthermore, $h^*(z)$ is a strictly increasing function of z, since the integrand of (3.3) for $0 \leqq z \leqq 1$ and of (3.5) for $z \geqq 1$ is strictly monotone for each value

of w, $0 < w < 1$. Consequently, $H(r')$ is a strictly increasing function of r' and $-1/(p-1) \leqq h^*(z) = H(r') \leqq 1$, which is the range of ρ.

As remarked before, $h^*(z) = H(r')$ is the unique minimum variance unbiased estimator of ρ.

We will now obtain the asymptotic distribution of $h^*(z)$. Note that z is distributed as

$$\frac{1 + (p-1)\rho}{1-\rho} \frac{a}{b(p-1)} F_{a,b},$$

and that $\sqrt{(p-1)N/p}\,(F_{a,b} - 1)$ is asymptotically $N(0, 1)$. Therefore, letting $z_0 = [1 + (p-1)\rho]/[(p-1)^2(1-\rho)]$, the quantity,

$$\sqrt{\frac{(p-1)N}{p}} \left[z \frac{(p-1)^2(1-\rho)}{1 + (p-1)\rho} - 1 \right] = \sqrt{\frac{N}{p}} \frac{(p-1)^{5/2}(1-\rho)}{1 + (p-1)\rho} (z - z_0)$$

is asymptotically $N(0, 1)$. But, by (3.3), denoting $N^{-5/6}$ by ϵ, we have, for $z \leqq 1$,

$$\begin{aligned} 1 - h^*(z) &= \frac{p}{p-1} \frac{N}{2} \int_0^\epsilon [1 - w - (p-1)zw]^{N/2}\, dw \\ &\qquad \cdot [1 + O(\epsilon^2)]^{N/2}[1 + O(\epsilon)] + NO(1-\epsilon)^{N/2} \\ &= \frac{p}{p-1} \frac{1}{1 + (p-1)z} + O\left(\frac{1}{N}\right), \end{aligned}$$

uniformly in z. We obtain the same result for $z \geqq 1$ from (3.4). Therefore

$$h^*(z) = \rho + (p-1)^2(1-\rho)^2(z - z_0)/p + O_p(1/N).$$

Therefore

$$\sqrt{N}\,[h^*(z) - \rho] = \sqrt{N}\,[H(r') - \rho] \quad \text{is asymptotically} \quad N(0, \sigma^2),$$

where $\sigma^2 = (1-\rho)^2 [1 + (p-1)\rho]^2/[p(p-1)]$.

Expanding r' about z_0 in (3.6) we find

$$r' = h^*(z) + O_p(1/N) = H(r') + O_p(1/N),$$

so that r' is asymptotically equivalent to $H(r')$. Incidentally, we find that $\sqrt{N}(r' - \rho)$ is asymptotically $N(0, \sigma^2)$, with the same σ^2.

In order to facilitate the use of the unbiased estimator in the bivariate case with n degrees of freedom, i.e., case (i) or (ii) with $p = 2$, Table 2 gives $H(r')$ and, (for easier interpolation), $H(r')/r'$ for $r' = 0(.1)1$ and $a = b = n = 2(2)30$. In this case, $H(r') = H_n(r')$ is an odd function of r'. The computation was carried out by means of the recursive relation

$$H_n(r') = \frac{n-2}{n-3}\left[\frac{r'}{1 - r'^2} - \frac{r'^2}{1 - r'^2} H_{n-2}(r')\right],$$

together with the initial conditions

$$H_2(r') = 1, \qquad H_4(r') = \frac{2r'}{1 + r'}, \qquad (r' > 0).$$

$H_n(r')$ was derived for $n \geqq 3$. For $n = 2$, the inversion in (3.1) must be carried out separately, and the result agrees with the final form of $h^*(z)$ in (3.4) and (3.6). The recursive relation is obtained by application of the relations [2, 2.8 (36) and (39)]. The same formulas give recursive relations for any values of p, a, b.

For the bivariate case with n degrees of freedom,

$$H(r') = r'F(\tfrac{1}{2}, 1; n/2; 1 - r'^2), \tag{3.8}$$

which is obtained from [2, 2.8 (36) and 2.11 (34)].

Approximations for $H(r')$ can be obtained from the expansion (2.2) applied to (3.8), which gives

$$\frac{H(r')}{r'} = 1 + \frac{1 - r'^2}{n} + \frac{3(1 - r'^2)^2}{n(n + 2)} + O(n^{-3}). \tag{3.9}$$

This gives $H(r')/r'$ within .01 for $n \geqq 19$ or .001 for $n \geqq 57$ if two terms are included, and within .01 for $n \geqq 12$ or .001 for $n \geqq 26$ if three terms are included. As in (2.6), the neglected terms in (3.9) are all positive and decreasing in r'^2 and n.

The k that minimizes the maximum over r of the absolute difference between $H(r')/r'$ and $1 + (1 - r'^2)/(n - k)$ is, for large n, $6(-1 + \sqrt{2}) = 2.48$. This suggests the approximation

$$\frac{H(r')}{r'} = 1 + \frac{1 - r'^2}{n - 5/2}. \tag{3.10}$$

This is accurate within .01 for $n \geqq 10$ or .001 for $n \geqq 26$.

4. Multiple correlation coefficient. Suppose we have N independent observations on a $p + 1$-variate normal distribution with mean vector μ and covariance matrix Σ, and we wish to give an unbiased estimator of the squared multiple correlation

$$\rho^2 = \rho^2_{0.(12\cdots p)} = 1 - \mathcal{R}/\mathcal{R}_{00},$$

where $\mathcal{R}$ is the determinant of the correlation matrix and $\mathcal{R}_{00}$ is its first cofactor. We are concerned with the cases (i) μ known, Σ unknown, and (ii) all parameters unknown.

As in 2.1, we confine ourselves to functions of

$$r^2 = r^2_{0.(12\cdots p)} = 1 - R/R_{00},$$

where R is the determinant of the appropriate (to (i) or (ii)) sample correlation matrix and R_{00} is its first cofactor.

The condition that $I(r^2)$ be unbiased is

$$\sum_{k=0}^{\infty} \frac{\Gamma^2\left(\frac{n}{2} + k\right)}{\Gamma\left(\frac{p}{2} + k\right)} \frac{\rho^{2k}}{k!} \int_0^1 I(r^2)(r^2)^{((p-2)/2)+k}(1 - r^2)^{(n-p-1)/2}\, dr^2$$

$$= \Gamma\left(\frac{n-p}{2}\right)\Gamma\left(\frac{n}{2}\right)(1 - \rho^2)^{-n/2}\rho^2,$$

where $n = N$ and $(N - 1)$ in cases (i) and (ii). Following the methods of Sec. 2, we obtain

$$I(r^2) = 1 - \frac{n-2}{n-p}(1 - r^2)F\left(1, 1; \frac{n-p+2}{2}; 1 - r^2\right).$$

As usual, $I(r^2)$ is strictly increasing in r^2, and differs from it only by terms of order $1/N$, and it is the unique minimum variance unbiased estimator of ρ^2. Also $I(1) = 1$. However, $I(0) = -p/(n - p - 2)$. We cannot hope for a non-negative unbiased estimator, since there is no region in the sample space having zero probability for $\rho^2 = 0$ and positive probability for $\rho^2 > 0$. For the same reason there can be no positive unbiased estimator of ρ either

REFERENCES

[1] Bateman Manuscript Project, California Institute of Technology, *Tables of Integral Transforms*, Vol 1, A. Erdelyi, editor, McGraw-Hill Book Co , New York, 1954.
[2] Bateman Manuscript Project, California Institute of Technology, *Higher Transcendental Functions*, Vol 1, A Erdelyi, editor, McGraw-Hill Book Co , New York, 1953
[3] H Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.
[4] G. Doetsch and Dietrich Voelker, *Die Zweidimensionale Laplace-Transformation*, Birkhauser Basel, Zurich, 1950
[5] D. B. DeLury, "Note on correlations," *Ann Math Stat.*, Vol 9 (1933), pp 149-151
[6] M. G. Kendall, *The Advanced Theory of Statistics*, Vol 1, Charles Griffin, London, 1947.

TRANSIENT ATOMIC MARKOV CHAINS WITH A DENUMERABLE NUMBER OF STATES[1]

By Leo Breiman

University of California

1. Introduction. Many of the more interesting transient Markov chains have the property that for any set of states A and any initial distribution, the probability of entering A infinitely often (i.o.) is either zero always or one always. This type of chain has been termed *atomic* by D. Blackwell [1] and is exemplified by the three-dimensional random walk or by the successive sums of independent, identically distributed random variables.

In this paper we investigate the "fine structure" of an atomic chain, that is, we try to characterize the class of all sets A such that $P(x_n \varepsilon A \text{ i.o.}) = 0$. The study is restricted to atomic chains with a countable set of states which, for convenience of notation, we identify with the integers, and with stationary transition probabilities $p_{ij}^{(n)}$.

The martingale convergence theorem is used in [1] to show that a necessary and sufficient condition for atomicity is that every bounded solution ϕ of

$$\phi(i) = \sum_j p_{ij}\phi(j)$$

be constant. We use as our main tool the semi-martingale convergence theorem and the corresponding equation $\phi(i) \geqq \sum_j p_{ij}\phi(j)$ and obtain a complete, but not simple, characterization of the fine structure of transient atomic chains.

To illustrate the use of the above characterization we prove two theorems regarding the return to equilibrium times $x_0, x_1, \cdots$ in the coin-tossing game. The latter of these is then used to prove that there exists no set of numbers $\{\lambda_m\}$ such that[2] $P(x_n \varepsilon A \text{ i.o.}) = 0 \Leftrightarrow \sum_{m\varepsilon A} \lambda_m < \infty$.

This last result shows that, in general, there is no simple resolution to the question of defining the fine structure. There are, however, a number of interesting transient atomic chains which have the property that every infinite set of states is entered infinitely often with probability one. These chains are the subject of papers by Chung and Derman [2], and Breiman [3].

2. Use of the semi-martingale theorem.

Theorem 1. *Let $x_0, x_1, \cdots$ be an atomic chain. Then for ϕ any nonnegative solution of*

(a) $$\phi(i) \geqq \sum_j p_{ij}\phi(j)$$

---

[1] This paper was prepared with the support of the Office of Ordnance Research, U.S. Army, under Contract DA-04-200-ORD-171.

[2] The referee has informed us that a similar theorem for the three-dimensional random walk has been proved by P. Erdös and B. J. Murdoch (unpublished).

which is finite for at least one value of i, $\phi(x_n)$ converges almost surely (a.s.) to a constant independent of the initial distribution.

PROOF. Let ϕ be a nonnegative solution of (a) with $\phi(0) < \infty$, and let R be the set of all states i such that P (entering $i \mid x_0 = 0) > 0$. From the atomicity and $P(x_n \varepsilon \bar{R} \text{ i.o.} \mid x_0) = 0$, where $\bar{R}$ is the complement of R, follows $P(x_n \varepsilon \bar{R}$ i.o.) $= 0$. Therefore, it is sufficient to prove the theorem for initial distributions concentrated on R noting that ϕ is finite on R. Pick any such distribution $\{p_j\}$ such that $\sum_j p_j\phi(j) < \infty$ and $p_j > 0$, all $j \varepsilon R$. The random variables $\phi(x_n)$ form a semi-martingale with respect to the fields generated by $x_0, x_1, \cdots$, since

$$E(\phi(x_n) \mid x_{n-1}, \cdots, x_0) = E(\phi(x_n) \mid x_{n-1}) = \sum p_{x_{n-1},j}\phi(j) \leqq \phi(x_{n-1}),$$

$$E|\phi(x_n)| = E\phi(x_n) \leqq E\phi(x_0) < \infty.$$

By the semi-martingale theorem [4], $\phi(x_n)$ converges a.s. Suppose this limit is nonconstant, then there will be a number $a > 0$ such that if A is the set of states defined by $\{j; \phi(j) \geqq a\}$, then $0 < P(x_n \varepsilon A$ i.o.) < 1. Hence the limit is constant, and since $\phi(x_n)$ must converge to this same constant with the initial distribution concentrated at any single state in R, the theorem is valid.

We note that the same result is true for ϕ any bounded solution of (a) because for any sufficiently large constant α, $\phi + \alpha$ is a positive solution.

A simple but informative corollary of the above theorem demonstrates the special applicability of (a) to the transient case.

COROLLARY. *All the states of an atomic chain are recurrent if and only if all bounded solutions of* (a) *are constant. For a transient atomic chain there is at least one nonconstant bounded solution of* (a)

PROOF. Let $\phi(i) = P$ (entering $i_0 \mid x_0 = i$), so that $\phi(i_0) = 1$, and

$$\phi(i) = E(P(\text{entering } i_0 \mid x_1, x_0) \mid x_0 = i) \geqq E(\phi(x_1) \mid x_0 = i) = \sum_j p_{ij}\phi(j).$$

If every solution of (a) is constant, then for every i_0 we have P(entering $i_0 \mid x_0 = i) = 1$. This implies that return to every state is certain Now assume that every state is recurrent and let ϕ be any bounded solution of (a). If $\phi(i) \neq \phi(j)$, $\phi(x_n)$ cannot possibly converge to a constant since both i and j are entered i.o. with probability one. If there are transient states present the function ϕ defined above cannot be constant for all i_0.

3. Characterization of the fine structure.

We use the notation

$$u_{ik} = E(\text{number of visits to } k \mid x_0 = i),$$

$$u_{ik}^{(n)} = \delta_{ik} + p_{ik} + \cdots + p_{ik}^{(n-1)} \qquad \delta_{ik} = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases}$$

and recall that $u_{ik} = \lim_n u_{ik}^{(n)}$.

THEOREM 2. *If $x_0, x_1, \cdots$ is an atomic chain, then for every nonnegative sequence of numbers $\{\alpha_k\}$ with $\sum_k u_{0k}\alpha_k < \infty$ and every $\epsilon > 0$, the set of states $A_\alpha = \{i; \sum_k u_{ik}\alpha_k \geqq \epsilon\}$ has the property $P(x_n \epsilon A_\alpha$ i.o.$) = 0$. Conversely, every*

TRANSIENT ATOMIC MARKOV CHAINS WITH A DENUMERABLE NUMBER OF STATES[1]

By Leo Breiman

University of California

1. Introduction. Many of the more interesting transient Markov chains have the property that for any set of states A and any initial distribution, the probability of entering A infinitely often (i.o.) is either zero always or one always. This type of chain has been termed *atomic* by D. Blackwell [1] and is exemplified by the three-dimensional random walk or by the successive sums of independent, identically distributed random variables.

In this paper we investigate the "fine structure" of an atomic chain, that is, we try to characterize the class of all sets A such that $P(x_n \varepsilon A \text{ i.o.}) = 0$. The study is restricted to atomic chains with a countable set of states which, for convenience of notation, we identify with the integers, and with stationary transition probabilities $p_{ij}^{(n)}$.

The martingale convergence theorem is used in [1] to show that a necessary and sufficient condition for atomicity is that every bounded solution ϕ of

$$\phi(i) = \sum_j p_{ij}\phi(j)$$

be constant. We use as our main tool the semi-martingale convergence theorem and the corresponding equation $\phi(i) \geqq \sum_j p_{ij}\phi(j)$ and obtain a complete, but not simple, characterization of the fine structure of transient atomic chains.

To illustrate the use of the above characterization we prove two theorems regarding the return to equilibrium times $x_0, x_1, \cdots$ in the coin-tossing game. The latter of these is then used to prove that there exists no set of numbers $\{\lambda_m\}$ such that[2] $P(x_n \varepsilon A \text{ i.o.}) = 0 \Leftrightarrow \sum_{m \varepsilon A} \lambda_m < \infty$.

This last result shows that, in general, there is no simple resolution to the question of defining the fine structure. There are, however, a number of interesting transient atomic chains which have the property that every infinite set of states is entered infinitely often with probability one. These chains are the subject of papers by Chung and Derman [2], and Breiman [3].

2. Use of the semi-martingale theorem.

Theorem 1. *Let $x_0, x_1, \cdots$ be an atomic chain. Then for ϕ any nonnegative solution of*

$$\text{(a)} \qquad \phi(i) \geqq \sum_j p_{ij}\phi(j)$$

[1] This paper was prepared with the support of the Office of Ordnance Research, U.S. Army, under Contract DA-04-200-ORD-171.

[2] The referee has informed us that a similar theorem for the three-dimensional random walk has been proved by P. Erdös and B. J. Murdoch (unpublished).

which is finite for at least one value of i, $\phi(x_n)$ converges almost surely (a.s.) to a constant independent of the initial distribution.

PROOF. Let ϕ be a nonnegative solution of (a) with $\phi(0) < \infty$, and let R be the set of all states i such that P (entering $i \mid x_0 = 0) > 0$. From the atomicity and $P(x_n \in \tilde{R}$ i.o. $\mid x_0) = 0$, where $\tilde{R}$ is the complement of R, follows $P(x_n \in \tilde{R}$ i.o.) $= 0$. Therefore, it is sufficient to prove the theorem for initial distributions concentrated on R noting that ϕ is finite on R. Pick any such distribution $\{p_j\}$ such that $\sum_j p_j\phi(j) < \infty$ and $p_j > 0$, all $j \in R$. The random variables $\phi(x_n)$ form a semi-martingale with respect to the fields generated by $x_0, x_1, \cdots$, since

$$E(\phi(x_n) \mid x_{n-1}, \cdots, x_0) = E(\phi(x_n) \mid x_{n-1}) = \sum p_{x_{n-1}j}\phi(j) \leqq \phi(x_{n-1}),$$

$$E|\phi(x_n)| = E\phi(x_n) \leqq E\phi(x_0) < \infty.$$

By the semi-martingale theorem [4], $\phi(x_n)$ converges a.s. Suppose this limit is nonconstant, then there will be a number $a > 0$ such that if A is the set of states defined by $\{j; \phi(j) \geqq a\}$, then $0 < P(x_n \in A$ i.o.$) < 1$. Hence the limit is constant, and since $\phi(x_n)$ must converge to this same constant with the initial distribution concentrated at any single state in R, the theorem is valid.

We note that the same result is true for ϕ any bounded solution of (a) because for any sufficiently large constant α, $\phi + \alpha$ is a positive solution.

A simple but informative corollary of the above theorem demonstrates the special applicability of (a) to the transient case.

COROLLARY. *All the states of an atomic chain are recurrent if and only if all bounded solutions of* (a) *are constant. For a transient atomic chain there is at least one nonconstant bounded solution of* (a)

PROOF. Let $\phi(i) = P$ (entering $i_0 \mid x_0 = i$), so that $\phi(i_0) = 1$, and

$$\phi(i) = E(P(\text{entering } i_0 \mid x_1, x_0) \mid x_0 = i) \geqq E(\phi(x_1) \mid x_0 = i) = \sum_j p_{ij}\phi(j).$$

If every solution of (a) is constant, then for every i_0 we have P(entering $i_0 \mid x_0 = i) = 1$. This implies that return to every state is certain. Now assume that every state is recurrent and let ϕ be any bounded solution of (a). If $\phi(i) \neq \phi(j)$, $\phi(x_n)$ cannot possibly converge to a constant since both i and j are entered i.o. with probability one. If there are transient states present the function ϕ defined above cannot be constant for all i_0.

3. Characterization of the fine structure. We use the notation

$$u_{ik} = E(\text{number of visits to } k \mid x_0 = i),$$

$$u_{ik}^{(n)} = \delta_{ik} + p_{ik} + \cdots + p_{ik}^{(n-1)} \qquad \delta_{ik} = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases}$$

and recall that $u_{ik} = \lim_n u_{ik}^{(n)}$.

THEOREM 2. *If $x_0, x_1, \cdots$ is an atomic chain, then for every nonnegative sequence of numbers $\{\alpha_k\}$ with $\sum_k u_{0k}\alpha_k < \infty$ and every $\epsilon > 0$, the set of states $A_\alpha = \{i; \sum_k u_{ik}\alpha_k \geqq \epsilon\}$ has the property $P(x_n \in A_\alpha$ i.o.$) = 0$. Conversely, every*

TRANSIENT ATOMIC MARKOV CHAINS WITH A DENUMERABLE NUMBER OF STATES[1]

By Leo Breiman

University of California

1. Introduction. Many of the more interesting transient Markov chains have the property that for any set of states A and any initial distribution, the probability of entering A infinitely often (i.o.) is either zero always or one always. This type of chain has been termed *atomic* by D. Blackwell [1] and is exemplified by the three-dimensional random walk or by the successive sums of independent, identically distributed random variables.

In this paper we investigate the "fine structure" of an atomic chain, that is, we try to characterize the class of all sets A such that $P(x_n \varepsilon A \text{ i.o.}) = 0$. The study is restricted to atomic chains with a countable set of states which, for convenience of notation, we identify with the integers, and with stationary transition probabilities $p_{ij}^{(n)}$.

The martingale convergence theorem is used in [1] to show that a necessary and sufficient condition for atomicity is that every bounded solution ϕ of

$$\phi(i) = \sum_j p_{ij}\phi(j)$$

be constant. We use as our main tool the semi-martingale convergence theorem and the corresponding equation $\phi(i) \geqq \sum_j p_{ij}\phi(j)$ and obtain a complete, but not simple, characterization of the fine structure of transient atomic chains.

To illustrate the use of the above characterization we prove two theorems regarding the return to equilibrium times $x_0, x_1, \cdots$ in the coin-tossing game. The latter of these is then used to prove that there exists no set of numbers $\{\lambda_m\}$ such that[2] $P(x_n \varepsilon A \text{ i.o.}) = 0 \Leftrightarrow \sum_{m \varepsilon A} \lambda_m < \infty$.

This last result shows that, in general, there is no simple resolution to the question of defining the fine structure. There are, however, a number of interesting transient atomic chains which have the property that every infinite set of states is entered infinitely often with probability one. These chains are the subject of papers by Chung and Derman [2], and Breiman [3].

2. Use of the semi-martingale theorem.

Theorem 1. *Let $x_0, x_1, \cdots$ be an atomic chain. Then for ϕ any nonnegative solution of*

(a) $$\phi(i) \geqq \sum_j p_{ij}\phi(j)$$

[1] This paper was prepared with the support of the Office of Ordnance Research, U.S. Army, under Contract DA-04-200-ORD-171.

[2] The referee has informed us that a similar theorem for the three-dimensional random walk has been proved by P. Erdös and B. J. Murdoch (unpublished).

which is finite for at least one value of i, $\phi(x_n)$ *converges almost surely (a.s.) to a constant independent of the initial distribution.*

Proof. Let ϕ be a nonnegative solution of (a) with $\phi(0) < \infty$, and let R be the set of all states i such that P (entering $i \mid x_0 = 0) > 0$. From the atomicity and $P(x_n \varepsilon \tilde{R} \text{ i.o.} \mid x_0) = 0$, where $\tilde{R}$ is the complement of R, follows $P(x_n \varepsilon \tilde{R}$ i.o.$) = 0$. Therefore, it is sufficient to prove the theorem for initial distributions concentrated on R noting that ϕ is finite on R. Pick any such distribution $\{p_j\}$ such that $\sum_j p_j\phi(j) < \infty$ and $p_j > 0$, all $j \varepsilon R$. The random variables $\phi(x_n)$ form a semi-martingale with respect to the fields generated by $x_0, x_1, \cdots$, since

$$E(\phi(x_n) \mid x_{n-1}, \cdots, x_0) = E(\phi(x_n) \mid x_{n-1}) = \sum p_{x_{n-1}j}\phi(j) \leqq \phi(x_{n-1}),$$

$$E|\phi(x_n)| = E\phi(x_n) \leqq E\phi(x_0) < \infty.$$

By the semi-martingale theorem [4], $\phi(x_n)$ converges a.s. Suppose this limit is nonconstant, then there will be a number $a > 0$ such that if A is the set of states defined by $\{j; \phi(j) \geqq a\}$, then $0 < P(x_n \varepsilon A \text{ i.o.}) < 1$. Hence the limit is constant, and since $\phi(x_n)$ must converge to this same constant with the initial distribution concentrated at any single state in R, the theorem is valid.

We note that the same result is true for ϕ any bounded solution of (a) because for any sufficiently large constant α, $\phi + \alpha$ is a positive solution.

A simple but informative corollary of the above theorem demonstrates the special applicability of (a) to the transient case.

Corollary. *All the states of an atomic chain are recurrent if and only if all bounded solutions of* (a) *are constant. For a transient atomic chain there is at least one nonconstant bounded solution of* (a)

Proof. Let $\phi(i) = P$ (entering $i_0 \mid x_0 = i$), so that $\phi(i_0) = 1$, and

$$\phi(i) = E(P(\text{entering } i_0 \mid x_1, x_0) \mid x_0 = i) \geqq E(\phi(x_1) \mid x_0 = i) = \sum_j p_{ij}\phi(j).$$

If every solution of (a) is constant, then for every i_0 we have $P(\text{entering } i_0 \mid x_0 = i) = 1$. This implies that return to every state is certain. Now assume that every state is recurrent and let ϕ be any bounded solution of (a). If $\phi(i) \neq \phi(j)$, $\phi(x_n)$ cannot possibly converge to a constant since both i and j are entered i.o. with probability one. If there are transient states present the function ϕ defined above cannot be constant for all i_0.

3. Characterization of the fine structure. We use the notation

$$u_{ik} = E(\text{number of visits to } k \mid x_0 = i),$$

$$u_{ik}^{(n)} = \delta_{ik} + p_{ik} + \cdots + p_{ik}^{(n-1)} \qquad \delta_{ik} = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases}$$

and recall that $u_{ik} = \lim_n u_{ik}^{(n)}$.

Theorem 2. *If* $x_0, x_1, \cdots$ *is an atomic chain, then for every nonnegative sequence of numbers* $\{\alpha_k\}$ *with* $\sum_k u_{0k}\alpha_k < \infty$ *and every* $\epsilon > 0$, *the set of states* $A_\alpha = \{i; \sum_k u_{ik}\alpha_k \geqq \epsilon\}$ *has the property* $P(x_n \epsilon A_\alpha \text{ i.o.}) = 0$. *Conversely, every*

set of states A such that $P(x_n \varepsilon A$ i.o.$) = 0$ is included in at least one of the sets A_α as defined above.

PROOF. Let $\{\alpha_k\}$ be a sequence fulfilling the conditions of the theorem and let $\phi(i) = \sum_k u_{ik}\alpha_k$. The identity $\sum_j p_{ij}u_{jk} = u_{ik} - \delta_{ik}$ leads to the equation $\Sigma_j p_{ij}\phi(j) = \phi(i) - \alpha_i$. Thus, theorem 1 applies and $\phi(x_n)$ converges a.s. to a constant. Since the properties in which we are interested do not, in an atomic chain, depend on initial conditions, it is sufficient to take $x_0 = 0$. Then, iterating the equation which ϕ satisfies,

$$E(\phi(x_n) \mid x_0 = 0) = \sum_k (u_{0k} - u_{0k}^{(n)})\alpha_k \to 0$$

and by a semi-martingale inequality ([4], p. 325) which states that

$$E(\text{a.s. limit}) \leqq E\phi(x_n)$$

we are able to conclude that the a.s. limit of $\phi(x_n)$ is identically zero. This implies that $P(\phi(x_n) \geqq \epsilon \text{ i.o.}) = 0$ and proves one part of the theorem.

To get the second part, let A be any set of states with $P(x_n \varepsilon A \text{ i.o.}) = 0$. Form the function $\phi(i) = P(\text{entering } A \mid x_0 = i)$, so that $\phi(i) = 1$, all $i \varepsilon A$. It is easy to verify that ϕ satisfies (a), and thus $\phi(x_n)$ converges a.s. to some constant. We deduce that this constant is zero by noting that $P(\text{entering } A$ after $n - 1$ steps$) = E\phi(x_n)$. Since $P(x_n \varepsilon A \text{ i.o.}) = 0$ we conclude that $E\phi(x_n) \to 0$ and apply the bounded convergence theorem to get the result. Let the nonnegative sequence $\{\alpha_i\}$ be defined by $\phi(i) = \alpha_i + \sum_j p_{ij}\phi(j)$. Iterating this equation

$$\phi(i) = \sum_j p_{ij}^{(n)}\phi(j) + \sum_j u_{ij}^{(n)}\alpha_j.$$

By the boundedness of ϕ the second sum converges to $\sum_j u_{ij}\alpha_j$. The first sum must also converge to some bounded limit sequence $\{\lambda(i)\}$. Since

$$\lambda(i) = \sum_j p_{ij}\lambda(j),$$

by Blackwell's theorem as quoted above this sequence is constant, and by the convergence of $\phi(x_n)$ to zero, $\lambda(i) \equiv 0$. The set A is contained in the set $A_\alpha = \{i; \sum_k u_{ik}\alpha_k \geqq 1\}$ which proves the theorem.

4. Two theorems concerning the coin-tossing game. We apply theorem 2 to the Markov chain $x_0, x_1, \cdots$ whose values are the successive times of return to equilibrium in the fair coin-tossing game. The set of states is the set of all nonnegative even integers and we use the fact that this chain, being the sum of independent and identically distributed random variables, is atomic. It is well known that

$$u_{ik} = 0, \qquad k < i,$$
$$\sim \frac{c}{\sqrt{(k-i)}}, \qquad k \gg i.$$

As it is evident that the characterization given in theorem 2 is invariant under asymptotic equivalence, we use $1/\sqrt{k-i}$ throughout this section in place of u_{ik} with the convention $\sqrt{0} = 1$ and $1/\sqrt{-} = 0$.

The first theorem we prove is similar to a theorem stated by Chung and Erdös [5].

THEOREM 3. *Let the sequence of even positive integers* $\{m_i\}$ *be such that the sequence* $\{\Delta_i\}$ *defined by* $\Delta_i = m_i - m_{i-1}$ *is nondecreasing. Then*

$$P(x_n \varepsilon \{m_i\} \text{ i.o.}) = 0 \Leftrightarrow \sum_i \frac{1}{\sqrt{m_i}} < \infty.$$

PROOF. If $\sum_i 1/\sqrt{m_i} < \infty$, the assertion follows immediately from the Borel-Cantelli lemma. Now assume that $P(x_n \varepsilon \{m_i\} \text{ i.o.}) = 0$, but that $\sum_i 1/\sqrt{m_i} = \infty$. By theorem 2, there is a nonnegative sequence $\{\alpha_i\}$ such that $\sum_k \alpha_k/\sqrt{k} < \infty$ and $\{m_i\} \subset \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \epsilon\}$. From this we have for all m_i

$$\sum_{k \geqq m_i} \frac{\alpha_k}{\sqrt{k - m_i}} \geqq \epsilon.$$

Define $\lambda_k^{(N)}$ by

$$\lambda_k^{(N)} = \frac{\displaystyle\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}\sqrt{k-m_i}}}{\displaystyle\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}}}.$$

It is evident that $\sum_k \lambda_k^{(N)}\alpha_k \geqq \epsilon$, *all* N, *and that* $\lim_N \lambda_k^{(N)} = 0$ *for* k *fixed.* We will show that $\lambda_k^{(N)} < c/\sqrt{k}$, all k, N, and conclude from the bounded convergence theorem the contradiction that $\lim_N \sum_k \lambda_k^{(N)}\alpha_k = 0$. To begin with, assume that $k \geqq m_N$, then

$$\sqrt{k}\,\lambda_k^{(N)} \leqq \frac{\displaystyle\sqrt{m_N}\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}\sqrt{m_N - m_i}}}{\displaystyle\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}}}.$$

By splitting the top sum into the two parts $m_i \leqq m_N/2$, $m_i > m_N/2$ and using our assumption concerning Δ_i, we get $\sqrt{k}\lambda_k^{(N)} \leqq 4$. Now if $k \leqq m_N$, let m_n be the largest of the m_i which is $\leqq k$. With this

$$\sqrt{k}\,\lambda_k^{(N)} \leqq \frac{\displaystyle\sqrt{m_n}\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}\sqrt{m_n - m_i}}}{\displaystyle\sum_{i=0}^{N} \frac{1}{\sqrt{m_i}}}$$

and repeating the above argument results again in $\sqrt{k}\lambda_k^{(N)} \leqq 4$.

It is clear that in the above context, a little more attention to the appropriate inequalities would result in a considerable weakening of the growth condition on the sequence $\{m_i\}$.

We can get a result in another direction by combining our characterization with different inequalities. Let all the states between and including n_1 and n_2, $n_2 \geqq n_1$, be called an interval and denoted by $[n_1, n_2]$.

set of states A such that $P(x_n \varepsilon A$ i.o.$) = 0$ is included in at least one of the sets A_α as defined above.

PROOF. Let $\{\alpha_k\}$ be a sequence fulfilling the conditions of the theorem and let $\phi(i) = \sum_k u_{ik}\alpha_k$. The identity $\sum_j p_{ij}u_{jk} = u_{ik} - \delta_{ik}$ leads to the equation $\Sigma_j p_{ij}\phi(j) = \phi(i) - \alpha_i$. Thus, theorem 1 applies and $\phi(x_n)$ converges a.s. to a constant. Since the properties in which we are interested do not, in an atomic chain, depend on initial conditions, it is sufficient to take $x_0 = 0$. Then, iterating the equation which ϕ satisfies,

$$E(\phi(x_n) \mid x_0 = 0) = \sum_k (u_{0k} - u_{0k}^{(n)})\alpha_k \to 0$$

and by a semi-martingale inequality ([4], p. 325) which states that

$$E(\text{a.s. limit}) \leqq E\phi(x_n)$$

we are able to conclude that the a.s. limit of $\phi(x_n)$ is identically zero. This implies that $P(\phi(x_n) \geqq \epsilon \text{ i.o.}) = 0$ and proves one part of the theorem.

To get the second part, let A be any set of states with $P(x_n \varepsilon A \text{ i.o.}) = 0$. Form the function $\phi(i) = P(\text{entering } A \mid x_0 = i)$, so that $\phi(i) = 1$, all $i \varepsilon A$. It is easy to verify that ϕ satisfies (a), and thus $\phi(x_n)$ converges a.s. to some constant. We deduce that this constant is zero by noting that P(entering A after $n - 1$ steps) $= E\phi(x_n)$. Since $P(x_n \varepsilon A \text{ i.o.}) = 0$ we conclude that $E\phi(x_n) \to 0$ and apply the bounded convergence theorem to get the result. Let the nonnegative sequence $\{\alpha_i\}$ be defined by $\phi(i) = \alpha_i + \sum_j p_{ij}\phi(j)$. Iterating this equation

$$\phi(i) = \sum_j p_{ij}^{(n)}\phi(j) + \sum_j u_{ij}^{(n)}\alpha_j.$$

By the boundedness of ϕ the second sum converges to $\sum_j u_{ij}\alpha_j$. The first sum must also converge to some bounded limit sequence $\{\lambda(i)\}$. Since

$$\lambda(i) = \textstyle\sum_j p_{ij}\lambda(j),$$

by Blackwell's theorem as quoted above this sequence is constant, and by the convergence of $\phi(x_n)$ to zero, $\lambda(i) \equiv 0$. The set A is contained in the set $A_\alpha = \{i; \sum_k u_{ik}\alpha_k \geqq 1\}$ which proves the theorem.

4. Two theorems concerning the coin-tossing game. We apply theorem 2 to the Markov chain $x_0, x_1, \cdots$ whose values are the successive times of return to equilibrium in the fair coin-tossing game. The set of states is the set of all nonnegative even integers and we use the fact that this chain, being the sum of independent and identically distributed random variables, is atomic. It is well known that

$$\begin{aligned} u_{ik} &= 0, && k < i, \\ &\sim \frac{c}{\sqrt{(k - i)}}, && k \gg i. \end{aligned}$$

As it is evident that the characterization given in theorem 2 is invariant under asymptotic equivalence, we use $1/\sqrt{k - i}$ throughout this section in place of u_{ik} with the convention $\sqrt{0} = 1$ and $1/\sqrt{-} = 0$.

The first theorem we prove is similar to a theorem stated by Chung and Erdös [5].

THEOREM 3. *Let the sequence of even positive integers* $\{m_\nu\}$ *be such that the sequence* $\{\Delta_\nu\}$ *defined by* $\Delta_\nu = m_\nu - m_{\nu-1}$ *is nondecreasing. Then*

$$P(x_n \in \{m_\nu\} \text{ i.o.}) = 0 \Leftrightarrow \sum_\nu \frac{1}{\sqrt{m_\nu}} < \infty.$$

PROOF. If $\sum_\nu 1/\sqrt{m_\nu} < \infty$, the assertion follows immediately from the Borel-Cantelli lemma. Now assume that $P(x_n \in \{m_\nu\} \text{ i.o.}) = 0$, but that $\sum_\nu 1/\sqrt{m_\nu} = \infty$. By theorem 2, there is a nonnegative sequence $\{\alpha_\nu\}$ such that $\sum_k \alpha_k/\sqrt{k} < \infty$ and $\{m_\nu\} \subset \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \epsilon\}$. From this we have for all m_ν

$$\sum_{k \geqq m_\nu} \frac{\alpha_k}{\sqrt{k - m_\nu}} \geqq \epsilon.$$

Define $\lambda_k^{(N)}$ by

$$\lambda_k^{(N)} = \frac{\sum_{\nu=0}^{N} \frac{1}{\sqrt{m_\nu}\sqrt{k - m_\nu}}}{\sum_{\nu=0}^{N} \frac{1}{\sqrt{m_\nu}}}.$$

It is evident that $\sum_k \lambda_k^{(N)} \alpha_k \geqq \epsilon$, all N, and that $\lim_N \lambda_k^{(N)} = 0$ for k fixed. We will show that $\lambda_k^{(N)} < c/\sqrt{k}$, all k, N, and conclude from the bounded convergence theorem the contradiction that $\lim_N \sum_k \lambda_k^{(N)} \alpha_k = 0$. To begin with, assume that $k \geqq m_N$, then

$$\sqrt{k}\,\lambda_k^{(N)} \leqq \frac{\sqrt{m_N} \sum_{\nu=0}^{N} \frac{1}{\sqrt{m_\nu}\sqrt{m_N - m_\nu}}}{\sum_{\nu=0}^{N} \frac{1}{\sqrt{m_\nu}}}.$$

By splitting the top sum into the two parts $m_\nu \leqq m_N/2$, $m_\nu > m_N/2$ and using our assumption concerning Δ_ν, we get $\sqrt{k}\lambda_k^{(N)} \leqq 4$. Now if $k \leqq m_N$, let m_n be the largest of the m_ν which is $\leqq k$. With this

$$\sqrt{k}\,\lambda_k^{(N)} \leqq \frac{\sqrt{m_n} \sum_{\nu=0}^{N} \frac{1}{\sqrt{m_\nu}\sqrt{m_n - m_\nu}}}{\sum_{\nu=0}^{N} \frac{1}{\sqrt{m_\nu}}}$$

and repeating the above argument results again in $\sqrt{k}\lambda_k^{(N)} \leqq 4$.

It is clear that in the above context, a little more attention to the appropriate inequalities would result in a considerable weakening of the growth condition on the sequence $\{m_\nu\}$.

We can get a result in another direction by combining our characterization with different inequalities. Let all the states between and including n_1 and n_2, $n_2 \geqq n_1$, be called an interval and denoted by $[n_1, n_2]$.

THEOREM 4. *If the sequence of disjoint finite intervals* $\{I_j\}$, $I_j = [m_j, M_j]$ *is such that for some* $\delta > 0$, $m_{j+1} \geqq (1 + \delta)M_j$, *then, denoting*

$$l_j = M_j - m_j + 2,$$

$$P(x_n \varepsilon \bigcup_j I_j \text{ i.o.}) = 0 \Leftrightarrow \sum_j \sqrt{l_j/M_j} < \infty.$$

PROOF. Define a sequence of intervals $I'_j = [m'_j, M'_j]$ by $m'_j = M_j$, $M'_j = M_j + \sqrt{l_j}$, where $\sqrt{l_j}$ is here to be interpreted as the greatest even integer less than $\sqrt{l_j}$. Let $\sum_j \sqrt{l_j/M_j} < \infty$ and define α_k by

$$\alpha_k = \begin{cases} 1 & \text{if } k \varepsilon \bigcup_j I'_j, \\ 0 & \text{otherwise.} \end{cases}$$

By these definitions

$$\sum_k \frac{\alpha_k}{\sqrt{k}} \leqq \frac{1}{2} \sum_j \sqrt{l_j/M_j} < \infty.$$

Thus the set $A = \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \frac{1}{2}\}$ has the property that $P(x_n \varepsilon A \textit{ i. o.}) = 0$. The set A includes, in particular, the integers i such that $i \leqq M_j$ and

$$\tfrac{1}{2} \leqq \sum_{k \varepsilon I'_j} \frac{1}{\sqrt{k - i}} \leqq \tfrac{1}{2}(\sqrt{M'_j - i} - \sqrt{m'_j - i}).$$

This inequality can be easily shown to be satisfied by all $i \geqq m_j$, which proves the theorem going one way.

To go the other way, assume that $P(x_n \epsilon \bigcup_j I_j \textit{ i. o.}) = 0$. Then there is a non-negative sequence a_k such that $\sum_k \alpha_k/\sqrt{k} < \infty$ and $\bigcup_j I_j \subset \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \epsilon\}$, from which, if $i \varepsilon I_j$, then $\sum_{k \geqq m_j} \alpha_k/\sqrt{k-i} \geqq \epsilon$. We wish to conclude that part of this sum is negligible and argue that if $i \varepsilon I_j$,and if j is sufficiently large

$$\sum_{k \geqq m_{j+1}} \sqrt{\frac{k}{k - i}} \frac{\alpha_k}{\sqrt{k}} \leqq \sqrt{\frac{m_{j+1}}{m_{j+1} - M_j}} \sum_{k \geqq m_{j+1}} \frac{\alpha_k}{\sqrt{k}} \leqq \frac{\epsilon}{2}$$

so that if $i \varepsilon I_j$,

$$\sum_{m_{j+1} \geqq k \geqq m_j} \frac{\alpha_k}{\sqrt{k - i}} \geqq \frac{\epsilon}{2}.$$

We sum this last inequality over $i \varepsilon I_j$ to get

$$\sum_{m_{j+1} > k \geqq m_j} \alpha_k \left(\sum_{i \varepsilon I_j} \frac{1}{\sqrt{k - i}} \right) \geqq \frac{\epsilon}{4} l_j.$$

It can be easily shown that

$$\sum_{i \varepsilon I_j} \frac{1}{\sqrt{k - i}} \leqq 4 \sqrt{\frac{M_j l_j}{k}},$$

and using this we conclude that

$$\sum_k \frac{\alpha_k}{\sqrt{k}} \geqq \frac{\epsilon}{16} \sum_j \sqrt{l_j/M_j}.$$

5. **The nonsimplicity of the fine structure of the coin-tossing game.** The purpose of this section is to prove the following theorem.

THEOREM 5. *Let $x_0, x_1, \cdots$ be the successive times of return to equilibrium in the fair coin-tossing game. Then there exists no weighting $\{\lambda_m\}$, $\lambda_m \geqq 0$ of the positive even integers such that*

$$P(x_n \varepsilon A \text{ i.o.}) = 0 \Leftrightarrow \sum_{m \varepsilon A} \lambda_m < \infty.$$

PROOF. Consider any set $\bigcup_j I_j$ where the I_j are disjoint finite intervals which we can represent as $[m_j, m_j(1 + \alpha_j)]$, $0 < \alpha_j \leqq 1$, where $m_{j+1} \geqq 3m_j$. By theorem 4

$$P(x_n \varepsilon \bigcup_j I_j \text{ i.o.}) = 0 \Leftrightarrow \sum_j \sqrt{\alpha_j} < \infty.$$

Let now $\{\lambda_m\}$ be any weighting of the positive even integers having the property stated in the theorem. By this property, $\lim_m \lambda_m = 0$ since otherwise we could find an indefinitely sparse set A which would be entered i.o. with probability one. We define a function $\phi(\alpha)$, $0 \leqq \alpha < \infty$ by

$$\phi(\alpha) = \liminf_n \sum_{m=n}^{ne^\alpha} \lambda_m,$$

where in writing the upper limit of summation as ne^α it is immaterial whether we take the next greater integer, or the previous integer.

PROPOSITION. *$\phi(\alpha)$ is monotone nondecreasing, $\phi(\alpha + \beta) \geqq \phi(\alpha) + \phi(\beta)$ and there is a neighborhood of the origin in which $\phi(\alpha) < \infty$.*

PROOF. The first assertion is immediate. As to the second, we write:

$$\liminf_n \left(\sum_{m=n}^{ne^\alpha e^\beta} \lambda_m \right) = \liminf_n \left(\sum_{m=n}^{ne^\alpha} \lambda_m + \sum_{m=ne^\alpha}^{ne^\alpha e^\beta} \lambda_m \right)$$
$$\geqq \liminf_n \left(\sum_{m=n}^{ne^\alpha} \lambda_m \right) + \liminf_n \left(\sum_{m=n}^{ne^\beta} \lambda_m \right).$$

Finally, suppose that $\phi(\alpha) = \infty$ for all $\alpha > 0$, and consider any sequence $\{\alpha_j\}$, $0 < \alpha_j \leqq 1$, such that $\sum_j \sqrt{\alpha_j} < \infty$. Since $\lim_n \sum_{m=n}^{n(1+\alpha_j)} \lambda_m = \infty$ for all j, we find a sequence of intervals $I_j = [m_j, m_j(1 + \alpha_j)]$ as far apart as desired having the undesirable property $\sum_j \sum_{m \varepsilon I_j} \lambda_m = \infty$.

To complete the proof of the theorem, we note that as a well-known consequence of the proposition there is a neighborhood N of the origin and a constant $q < \infty$ such that $\phi(\alpha) \leqq q\alpha$, $\alpha \varepsilon N$. Take $\{\alpha_j\}$, $\alpha_j > 0$, such that $\sum_j \alpha_j < \infty$ but $\sum_j \sqrt{\alpha_j} = \infty$, and $\{\alpha_j\} \subset N$. Then we may find a sequence $\{m_j\}$ increasing as rapidly as desired such that

$$\sum_{m=m_j}^{m_j(1+\alpha_j)} \lambda_m \leqq 2q\alpha_j.$$

Hence, taking $I_j = [m_j, m_j(1 + \alpha_j)]$ we have

$$\sum_j \sqrt{\alpha_j} = \infty \quad \text{but} \quad \sum_j \sum_{m \varepsilon I_j} \lambda_m < \infty.$$

THEOREM 4. *If the sequence of disjoint finite intervals* $\{I_j\}$, $I_j = [m_j, M_j]$ *is such that for some* $\delta > 0$, $m_{j+1} \geqq (1 + \delta)M_j$, *then, denoting*

$$l_j = M_j - m_j + 2,$$

$$P(x_n \varepsilon \bigcup_j I_j \text{ i.o.}) = 0 \Leftrightarrow \sum_j \sqrt{l_j/M_j} < \infty.$$

PROOF. Define a sequence of intervals $I_j' = [m_j', M_j']$ by $m_j' = M_j$, $M_j' = M_j + \sqrt{l_j}$, where $\sqrt{l_j}$ is here to be interpreted as the greatest even integer less than $\sqrt{l_j}$. Let $\sum_j \sqrt{l_j/M_j} < \infty$ and define α_k by

$$\alpha_k = \begin{cases} 1 & \text{if } k \varepsilon \bigcup_j I_j', \\ 0 & \text{otherwise.} \end{cases}$$

By these definitions

$$\sum_k \frac{\alpha_k}{\sqrt{k}} \leqq \frac{1}{2} \sum_j \sqrt{l_j/M_j} < \infty.$$

Thus the set $A = \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \frac{1}{2}\}$ has the property that $P(x_n \varepsilon A \text{ i.o.}) = 0$. The set A includes, in particular, the integers i such that $i \leqq M_j$ and

$$\tfrac{1}{2} \leqq \sum_{k \varepsilon I_j'} \frac{1}{\sqrt{k - i}} \leqq \tfrac{1}{2}(\sqrt{M_j' - i} - \sqrt{m_j' - i}).$$

This inequality can be easily shown to be satisfied by all $i \geqq m_j$, which proves the theorem going one way.

To go the other way, assume that $P(x_n \varepsilon \bigcup_j I_j \text{ i.o.}) = 0$. Then there is a non-negative sequence α_k such that $\sum_k \alpha_k/\sqrt{k} < \infty$ and $\bigcup_j I_j \subset \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \epsilon\}$, from which, if $i \varepsilon I_j$, then $\sum_{k \geqq m_j} \alpha_k/\sqrt{k-i} \geqq \epsilon$. We wish to conclude that part of this sum is negligible and argue that if $i \varepsilon I_j$, and if j is sufficiently large

$$\sum_{k \geqq m_{j+1}} \sqrt{\frac{k}{k - i}} \frac{\alpha_k}{\sqrt{k}} \leqq \sqrt{\frac{m_{j+1}}{m_{j+1} - M_j}} \sum_{k \geqq m_{j+1}} \frac{\alpha_k}{\sqrt{k}} \leqq \frac{\epsilon}{2}$$

so that if $i \varepsilon I_j$,

$$\sum_{m_{j+1} > k \geqq m_j} \frac{\alpha_k}{\sqrt{k - i}} \geqq \frac{\epsilon}{2}.$$

We sum this last inequality over $i \varepsilon I_j$ to get

$$\sum_{m_{j+1} > k \geqq m_j} \alpha_k \left(\sum_{i \varepsilon I_j} \frac{1}{\sqrt{k - i}} \right) \geqq \frac{\epsilon}{4} l_j.$$

It can be easily shown that

$$\sum_{i \varepsilon I_j} \frac{1}{\sqrt{k - i}} \leqq 4 \sqrt{\frac{M_j l_j}{k}},$$

and using this we conclude that

$$\sum_k \frac{\alpha_k}{\sqrt{k}} \geqq \frac{\epsilon}{16} \sum_j \sqrt{l_j/M_j}.$$

5. The nonsimplicity of the fine structure of the coin-tossing game. The purpose of this section is to prove the following theorem.

THEOREM 5. *Let $x_0, x_1, \cdots$ be the successive times of return to equilibrium in the fair coin-tossing game. Then there exists no weighting $\{\lambda_m\}$, $\lambda_m \geqq 0$ of the positive even integers such that*

$$P(x_n \varepsilon A \text{ i.o.}) = 0 \Leftrightarrow \sum_{m \varepsilon A} \lambda_m < \infty.$$

PROOF. Consider any set $\bigcup_j I_j$ where the I_j are disjoint finite intervals which we can represent as $[m_j, m_j(1 + \alpha_j)]$, $0 < \alpha_j \leqq 1$, where $m_{j+1} \geqq 3m_j$. By theorem 4

$$P(x_n \varepsilon \bigcup_j I_j \text{ i.o}) = 0 \Leftrightarrow \sum_j \sqrt{\alpha_j} < \infty$$

Let now $\{\lambda_m\}$ be any weighting of the positive even integers having the property stated in the theorem. By this property, $\lim_m \lambda_m = 0$ since otherwise we could find an indefinitely sparse set A which would be entered i.o. with probability one. We define a function $\phi(\alpha)$, $0 \leqq \alpha < \infty$ by

$$\phi(\alpha) = \liminf_n \sum_{m=n}^{ne^\alpha} \lambda_m,$$

where in writing the upper limit of summation as ne^α it is immaterial whether we take the next greater integer, or the previous integer.

PROPOSITION. *$\phi(\alpha)$ is monotone nondecreasing, $\phi(\alpha + \beta) \geqq \phi(\alpha) + \phi(\beta)$ and there is a neighborhood of the origin in which $\phi(\alpha) < \infty$.*

PROOF. The first assertion is immediate. As to the second, we write:

$$\liminf_n \left(\sum_{m=n}^{ne^\alpha e^\beta} \lambda_m \right) = \liminf_n \left(\sum_{m=n}^{ne^\alpha} \lambda_m + \sum_{m=ne^\alpha}^{ne^\alpha e^\beta} \lambda_m \right)$$

$$\geqq \liminf_n \left(\sum_{m=n}^{ne^\alpha} \lambda_m \right) + \liminf_n \left(\sum_{m=n}^{ne^\beta} \lambda_m \right).$$

Finally, suppose that $\phi(\alpha) = \infty$ for all $\alpha > 0$, and consider any sequence $\{\alpha_j\}$, $0 < \alpha_j \leqq 1$, such that $\sum_j \sqrt{\alpha_j} < \infty$. Since $\lim_n \sum_{m=n}^{n(1+\alpha_j)} \lambda_m = \infty$ for all j, we find a sequence of intervals $I_j = [m_j, m_j(1 + \alpha_j)]$ as far apart as desired having the undesirable property $\sum_j \sum_{m \varepsilon I_j} \lambda_m = \infty$

To complete the proof of the theorem, we note that as a well-known consequence of the proposition there is a neighborhood N of the origin and a constant $q < \infty$ such that $\phi(\alpha) \leqq q\alpha$, $\alpha \varepsilon N$. Take $\{\alpha_j\}$, $\alpha_j > 0$, such that $\sum_j \alpha_j < \infty$ but $\sum_j \sqrt{\alpha_j} = \infty$, and $\{\alpha_j\} \subset N$. Then we may find a sequence $\{m_j\}$ increasing as rapidly as desired such that

$$\sum_{m=m_j}^{m_j(1+\alpha_j)} \lambda_m \leqq 2q\alpha_j.$$

Hence, taking $I_j = [m_j, m_j(1 + \alpha_j)]$ we have

$$\sum_j \sqrt{\alpha_j} = \infty \qquad \text{but} \qquad \sum_j \sum_{m \varepsilon I_j} \lambda_m < \infty.$$

THEOREM 4. *If the sequence of disjoint finite intervals* $\{I_j\}$, $I_j = [m_j, M_j]$ *is such that for some* $\delta > 0$, $m_{j+1} \geqq (1 + \delta)M_j$, *then, denoting*

$$l_j = M_j - m_j + 2,$$
$$P(x_n \varepsilon \bigcup_j I_j \text{ i.o.}) = 0 \Leftrightarrow \sum_j \sqrt{l_j/M_j} < \infty.$$

PROOF. Define a sequence of intervals $I'_j = [m'_j, M'_j]$ by $m'_j = M_j$, $M'_j = M_j + \sqrt{l_j}$, where $\sqrt{l_j}$ is here to be interpreted as the greatest even integer less than $\sqrt{l_j}$. Let $\sum_j \sqrt{l_j/M_j} < \infty$ and define α_k by

$$\alpha_k = \begin{cases} 1 & \text{if } k \varepsilon \bigcup_j I'_j, \\ 0 & \text{otherwise.} \end{cases}$$

By these definitions

$$\sum_k \frac{\alpha_k}{\sqrt{k}} \leqq \frac{1}{2} \sum_j \sqrt{l_j/M_j} < \infty.$$

Thus the set $A = \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \frac{1}{2}\}$ has the property that $P(x_n \varepsilon A \ i.o.) = 0$. The set A includes, in particular, the integers i such that $i \leqq M_j$ and

$$\tfrac{1}{2} \leqq \sum_{k \varepsilon I'_j} \frac{1}{\sqrt{k - i}} \leqq \tfrac{1}{2}(\sqrt{M'_j - i} - \sqrt{m'_j - i}).$$

This inequality can be easily shown to be satisfied by all $i \geqq m_j$, which proves the theorem going one way.

To go the other way, assume that $P(x_n \epsilon \bigcup_j I_j \ i.\ o.) = 0$. Then there is a non-negative sequence a_k such that $\sum_k \alpha_k/\sqrt{k} < \infty$ and $\bigcup_j I_j \subset \{i; \sum_k \alpha_k/\sqrt{k-i} \geqq \epsilon\}$, from which, if $i \varepsilon I_j$, then $\sum_{k \geqq m_j} \alpha_k/\sqrt{k-i} \geqq \epsilon$. We wish to conclude that part of this sum is negligible and argue that if $i \varepsilon I_j$, and if j is sufficiently large

$$\sum_{k \geqq m_{j+1}} \sqrt{\frac{k}{k - i}} \frac{\alpha_k}{\sqrt{k}} \leqq \sqrt{\frac{m_{j+1}}{m_{j+1} - M_j}} \sum_{k \geqq m_{j+1}} \frac{\alpha_k}{\sqrt{k}} \leqq \frac{\epsilon}{2}$$

so that if $i \varepsilon I_j$,

$$\sum_{m_{j+1} \geqq k \geqq m_j} \frac{\alpha_k}{\sqrt{k - i}} \geqq \frac{\epsilon}{2}.$$

We sum this last inequality over $i \varepsilon I_j$ to get

$$\sum_{m_{j+1} > k \geqq m_j} \alpha_k \left(\sum_{i \varepsilon I_j} \frac{1}{\sqrt{k - i}} \right) \geqq \frac{\epsilon}{4} l_j.$$

It can be easily shown that

$$\sum_{i \varepsilon I_j} \frac{1}{\sqrt{k - i}} \leqq 4 \sqrt{\frac{M_j l_j}{k}},$$

and using this we conclude that

$$\sum_k \frac{\alpha_k}{\sqrt{k}} \geqq \frac{\epsilon}{16} \sum_j \sqrt{l_j/M_j}.$$

A CONSTRUCTION FOR ROOM'S SQUARES AND AN APPLICATION IN EXPERIMENTAL DESIGN

BY J. W. ARCHBOLD AND N. L. JOHNSON

University College, London

1. T. G. Room [1] recently proposed the following problem: To arrange the $n(2n - 1)$ symbols rs (which is the same as sr) formed from all pairs of $2n$ different digits in a square of $2n - 1$ rows and columns such that in each row and column there appear n symbols (and $n - 1$ blanks) which among them contain all $2n$ digits.

He remarked that the problem is soluble when $n = 1$ (trivially) and $n = 4$ but not when $n = 2$ or 3; and he gave one solution for $n = 4$.

Squares of such a type have uses in experimental designs. We explain below a simple construction for squares where n has the form 2^{2m-1}. Each square constructed in this way is represented in a canonical form by applying a well-known theorem of J. Singer [2]. In this form as soon as the top row of entries in a square is known, all the other entries may be written down immediately by means of a straight-forward cyclic process. Thus an index of first rows is all that is necessary to catalogue squares in their canonical forms.

It may be permissible to give here a slight modification of the proof of Singer's theorem in order to show a natural application of the regular representation of linear algebras.

2. Let $\mathfrak{A}$ be a linear associative algebra, of order m and with modulus, over a commutative field K. It is well known that $\mathfrak{A}$ is isomorphic with an algebra of $m \times m$ matrices whose elements belong to K (c f Macduffee [3], Section 123).

A Galois field $GF(p^{mn})$ is such a linear algebra over a $GF(p^n)$. If the elements of the $GF(p^{mn})$ are $0, \alpha, \alpha^2, \cdots, \alpha^{p^{mn}-1} = 1$ the irreducible equation, of degree m and with coefficients in $GF(p^n)$,

$$f(x) \equiv x^m - a_1x^{m-1} - \cdots - a_m = 0$$

which is satisfied by α is called primitive (Dickson [4], Section 35). A basis for the algebra consists of $1, \alpha, \alpha^2, \cdots, \alpha^{m-1}$ and the modulus is 1.

The primitive equation is both the minimum and characteristic equation of the companion matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ a_m & a_{m-1} & a_{m-2} & \cdots & a_1 \end{pmatrix}$$

Received December 20, 1956; revised October 16, 1957

It is a pleasure to acknowledge my debt to David Blackwell who brought my attention to the problems treated above.

REFERENCES

[1] D. Blackwell, "On transient Markov processes with a countable number of states and stationary transition probabilities," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 654-658.
[2] K. L. Chung and C. Derman, "Non-recurrent random walks," *Pacific J. Math.*, Vol. 6 (1956), pp. 441-447.
[3] L. Breiman, "On transient Markov chains with application to the uniqueness problem for Markov processes," *Ann. Math. Stat.*, to be published.
[4] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, 1953.
[5] K. L. Chung and P. Erdös, "On the application of the Borel-Cantelli lemma," *Trans. Amer. Math. Soc.*, Vol. 72 (1952), pp. 179-186.

A CONSTRUCTION FOR ROOM'S SQUARES AND AN APPLICATION IN EXPERIMENTAL DESIGN

By J. W. Archbold and N. L. Johnson

University College, London

1. T. G. Room [1] recently proposed the following problem: To arrange the $n(2n-1)$ symbols rs (which is the same as sr) formed from all pairs of $2n$ different digits in a square of $2n-1$ rows and columns such that in each row and column there appear n symbols (and $n-1$ blanks) which among them contain all $2n$ digits.

He remarked that the problem is soluble when $n = 1$ (trivially) and $n = 4$ but not when $n = 2$ or 3; and he gave one solution for $n = 4$.

Squares of such a type have uses in experimental designs. We explain below a simple construction for squares where n has the form 2^{2m-1}. Each square constructed in this way is represented in a canonical form by applying a well-known theorem of J. Singer [2]. In this form as soon as the top row of entries in a square is known, all the other entries may be written down immediately by means of a straight-forward cyclic process. Thus an index of first rows is all that is necessary to catalogue squares in their canonical forms.

It may be permissible to give here a slight modification of the proof of Singer's theorem in order to show a natural application of the regular representation of linear algebras.

2. Let $\mathfrak{A}$ be a linear associative algebra, of order m and with modulus, over a commutative field K. It is well known that $\mathfrak{A}$ is isomorphic with an algebra of $m \times m$ matrices whose elements belong to K (c f. Macduffee [3], Section 123).

A Galois field $GF(p^{mn})$ is such a linear algebra over a $GF(p^n)$. If the elements of the $GF(p^{mn})$ are $0, \alpha, \alpha^2, \cdots, \alpha^{p^{mn}-1} = 1$ the irreducible equation, of degree m and with coefficients in $GF(p^n)$,

$$f(x) \equiv x^m - a_1x^{m-1} - \cdots - a_m = 0$$

which is satisfied by α is called primitive (Dickson [4], Section 35) A basis for the algebra consists of $1, \alpha, \alpha^2, \cdots, \alpha^{m-1}$ and the modulus is 1.

The primitive equation is both the minimum and characteristic equation of the companion matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdot & 1 \\ a_m & a_{m-1} & a_{m-2} & \cdots & a_1 \end{pmatrix}$$

Received December 20, 1956; revised October 16, 1957.

The correspondence $\alpha^r \leftrightarrow \mathbf{A}^r$ determines an isomorphy, or regular representation, of the $GF(p^{mn})$ on the algebra or Galois field whose elements are the $m \times m$ matrices $0, \mathbf{A}, \mathbf{A}^2, \cdots, \mathbf{A}^{p^{mn}-1} = \mathbf{I}$, where $\mathbf{I}$ is the unit matrix (c.f. Macduffee [3], Section 109). If $N = 1 + p^n + \cdots + p^{(m-1)n}$ then the matrices $\mathbf{A}^{jN}$, for $j = 1, \cdots, p^n - 1$, are the multiples of $\mathbf{I}$ by the elements of $GF(p^n)$ and form a sub-algebra, of matrices, isomorphic with $GF(p^n)$.

In a finite projective space $PG(m-1, p^n)$ over the $GF(p^n)$, let $\mathbf{x}$ and $\mathbf{y}$ denote column coordinate vectors. Then *the equation* $k\mathbf{y} = \mathbf{A}\mathbf{x}$, *where* k *is any non-zero element of* $GF(p^n)$, *determines a homography in the space of period* N. This is Singer's theorem; and the proof differs from his more in form than substance. It is significant for us that N is also the number of points in the space.

3. Confine attention now to the case where $p = 2$ and $n = 1$. The space, a $PG\ (m-1, 2)$, contains $\mu, = 2^m - 1$, points, with three on every line.

The following are primitive irreducible polynomials over $GF(2)$:

$$\begin{array}{lll} x^2 - (x+1), & x^3 - (x+1), & x^4 - (x+1), \\ x^5 - (x^2+1), & x^6 - (x+1), & x^7 - (x+1) \\ x^8 - (x^4 + x^3 + x^2 + 1), & & x^9 - (x^8 + x^4 + x^3 + x^2 + 1). \end{array}$$

This list is taken from Dickson ([4], p. 44); it is not exhaustive for the degrees mentioned but for each degree the second largest exponent of x is as small as possible.

For a given m, choose any appropriate primitive polynomial and consider the associated homography of $PG\ (m-1, 2)$ of period μ. If P_1 is any point of the space, let its successive transforms under the homography be $P_2, P_3, \cdots, P_\mu\ (P_{\mu+1} = P_1)$.

Now consider the space $PG\ (m-1, 2)$ as being a prime in a $PG\ (m, 2)$. To achieve this, suppose $\mathbf{x}_1, \cdots, \mathbf{x}_\mu$ are coordinate vectors for $P_1, \cdots, P_\mu$. Then coordinate vectors for all but one of the points in $PG(m, 2)$ are obtained by adding a further zero or unit coordinate at the end of each $\mathbf{x}_i$; and the last point by taking coordinates consisting of m zeros followed by 1. Denote this last point by Q_0 and let Q_i be the third point on the line Q_0P_i; Q_i and P_i have the same first m coordinates.

To fix ideas, take $m = 3$ and $f(x) = x^3 - x - 1$. Then $\mu = 7$ and the corresponding homography is

$$k\begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$$

or

$$y_0 : y_1 : y_2 = x_1 : x_2 : x_0 + x_1 .$$

Starting with $x_0 = 1, x_1 = x_2 = 0$, we obtain for $PG(3, 2)$ the following points:

$$\begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 \\ \begin{pmatrix}1\\0\\0\\0\end{pmatrix} & \begin{pmatrix}0\\0\\1\\0\end{pmatrix} & \begin{pmatrix}0\\1\\0\\0\end{pmatrix} & \begin{pmatrix}1\\0\\1\\0\end{pmatrix} & \begin{pmatrix}0\\1\\1\\0\end{pmatrix} & \begin{pmatrix}1\\1\\1\\0\end{pmatrix} & \begin{pmatrix}1\\1\\0\\0\end{pmatrix} \end{matrix}$$

$$\begin{matrix} Q_0 & Q_1 & Q_2 & Q_3 & Q_4 & Q_5 & Q_6 & Q_7 \\ \begin{pmatrix}0\\0\\0\\1\end{pmatrix} & \begin{pmatrix}1\\0\\0\\1\end{pmatrix} & \begin{pmatrix}0\\0\\1\\1\end{pmatrix} & \begin{pmatrix}0\\1\\0\\1\end{pmatrix} & \begin{pmatrix}1\\0\\1\\1\end{pmatrix} & \begin{pmatrix}0\\1\\1\\1\end{pmatrix} & \begin{pmatrix}1\\1\\1\\1\end{pmatrix} & \begin{pmatrix}1\\1\\0\\1\end{pmatrix} \end{matrix}.$$

4. The idea is now to rename the points $Q_1, \cdots, Q_\mu$ as $R_1, \cdots, R_\mu$ in some order to be determined with the object, when possible, of ensuring that whenever the line Q_iQ_j passes through a point P_r then the line R_iR_j passes through a different point P_s.

The various incidences are then registered in a table of μ rows and μ columns as follows: if the line Q_iQ_j passes through P_r and R_iR_j passes through P_s make the entry

$$i, j \qquad (\text{or } j, i)$$

in the place belonging to the rth row and sth column of the table.

The number of entries in each row and column is the number of lines through a point of $PG(m, 2)$ which do not lie in a prime through the point. This number is $(2^m - 1) - (2^{m-1} - 1) = 2^{m-1}$. *And the entries in every row and column are* all the integers $0, 1, 2, 3, \cdots, 2^m - 1$ taken in pairs. No two pairs are the same and there are $2^{m-1}(2^m - 1)$ entries altogether.

In the cases examined below, the desired objective is reached when m is odd by defining R_t to be Q_u, where $u = 2^m - t$; and then no position in the incidence table contains more than one entry of the form (i, j). *When m is even, the same* definition is used for R_t but this leads to two entries in each position in the south-

TABLE 1

	1	2	3	4	5	6	7
1			24		56	37	01
2		35		67	41	02	
3	46		71	52	03		
4		12	63	04			57
5	23	74	05			61	
6	15	06			72		34
7	07			13		45	26

TABLE 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1					3, 9		12, 13			2, 5	7, 14		4, 15	8, 10 0, 2	0, 1 6, 11
2				4, 10		13, 14			3, 6	8, 15		5, 1	9, 11 0, 3	7, 12	
3			5, 11		14, 15			4, 7	9, 1		6, 2	10, 12 0, 4	8, 13		
4		6, 12		15, 1			5, 8	10, 2		7, 3	11, 13 0, 5	9, 14			
5	7, 13		1, 2			6, 9	11, 3		8, 4	12, 14 0, 6	10, 15				
6		2, 3			7, 10	12, 4		9, 5	13, 15 0, 7	11, 1					8, 14
7	3, 4			8, 11	13, 5		10, 6	14, 1 0, 8	12, 2					9, 15	
8			9, 12	14, 6		11, 7	15, 2 0, 9	13, 3					10, 1		4, 5
9		10, 13	15, 7		12, 8	1, 3 0, 10	14, 4					11, 2		5, 6	
10	11, 14	1, 8		13, 9	2, 4 0, 11	15, 5					12, 3		6, 7		
11	2, 9		14, 10	3, 5 0, 12	1, 6					13, 4		7, 8			12, 15
12		15, 11	4, 6 0, 13	2, 7					14, 5		8, 9			13, 1	3, 10
13	1, 12	5, 7 0, 14	3, 8					15, 6		9, 10			14, 2	4, 11	
14	6, 8 0, 15	4, 9					1, 7		10, 11			15, 3	5, 12		2, 13
15	5, 10					2, 8		11, 12			1, 4	6, 13		3, 14	7, 9

to north-east diagonal of the table: no better definition for R_i has been ›ed which will prevent two entries from occurring in the same position.

For the case $m = 3$, which we began to consider in Section 3, let us define › be Q_{8-i} ($i = 1, \cdots, 7$). We then obtain the incidences shown in Table 1. will be noticed that, beginning with the second, each row or column is ined by a cyclic change in the positions, and values modulo 7, of the entries ıe preceding row or column: that is, if X_{rs}, Y_{rs} are the entries in row r and ›nn s and $X_{rs} \neq 0$, then, modulo 7,

$$X_{r,s} \equiv 1 + X_{r-1,s+1}, \qquad Y_{r,s} \equiv 1 + Y_{r-1,s+1}.$$

ıe whole table is therefore completely determined by the entries in any one or column.

For $m = 4$, we have $\mu = 15$ and we take $f(x) = x^4 - x - 1$. R_i is now ıed to be Q_{16-i} for $i = 1, \cdots, 15$. Table 2 is obtained.

ere the NE–SW diagonal is shared by two sets of entries This is a character- feature arising when m is even but not when m is odd.

ı fact, going now to the simplest case where $m = 2$ and $f(x) = x^2 - x - 1$, table which arises is as follows:

	1	2	3
1			0 1 2 3
2		0 2 3 1	
3	0 3 1 2		

'or $m = 5$, $\mu = 31$. Take $f(x) = x^5 - x^2 - 1$. Define R_i to be Q_{32-i}. Then we ain Table 3 (only the first line of entries need be given).

TABLE 3

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
20	8,23	9,21	14,15		10,17		27,29				2,19			18,31

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
26	5,11						16,25	3,6			13,24		7,28	4,30	0,1

TABLE 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1					3, 9		12, 13			2, 5	7, 14		4, 15	8, 10 0, 2	0, 1 6, 11
2				4, 10		13, 14			3, 6	8, 15		5, 1	9, 11 0, 3	7, 12	
3			5, 11		14, 15			4, 7	9, 1		6, 2	10, 12 0, 4	8, 13		
4		6, 12		15, 1			5, 8	10, 2		7, 3	11, 13 0, 5	9, 14			
5	7, 13		1, 2			6, 9	11, 3		8, 4	12, 14 0, 6	10, 15				
6		2, 3			7, 10	12, 4		9, 5	13, 15 0, 7	11, 1					8, 14
7	3, 4			8, 11	13, 5		10, 6	14, 1 0, 8	12, 2					9, 15	
8			9, 12	14, 6		11, 7	15, 2 0, 9	13, 3					10, 1		4, 5
9		10, 13	15, 7		12, 8	1, 3 0, 10	14, 4					11, 2		5, 6	
10	11, 14	1, 8		13, 9	2, 4 0, 11	15, 5					12, 3		6, 7		
11	2, 9		14, 10	3, 5 0, 12	1, 6					13, 4		7, 8			12, 15
12		15, 11	4, 6 0, 13	2, 7					14, 5		8, 9			13, 1	3, 10
13	1, 12	5, 7 0, 14	3, 8					15, 6		9, 10			14, 2	4, 11	
14	6, 8 0, 15	4, 9					1, 7		10, 11				5, 12		2, 13
15	5, 10					2, 8		11, 12			1, 4	6, 13		3, 14	7, 9

west to north-east diagonal of the table: no better definition for R_i has been devised which will prevent two entries from occurring in the same position.

5. For the case $m = 3$, which we began to consider in Section 3, let us define R_i to be Q_{8-i} $(i = 1, \cdots, 7)$. We then obtain the incidences shown in Table 1. It will be noticed that, beginning with the second, each row or column is obtained by a cyclic change in the positions, and values modulo 7, of the entries in the preceding row or column: that is, if X_{rs}, Y_{rs} are the entries in row r and column s and $X_{rs} \neq 0$, then, modulo 7,

$$X_{r,s} \equiv 1 + X_{r-1,s+1}, \qquad Y_{r,s} \equiv 1 + Y_{r-1,s+1}.$$

The whole table is therefore completely determined by the entries in any one row or column.

6. For $m = 4$, we have $\mu = 15$ and we take $f(x) = x^4 - x - 1$. R_i is now defined to be Q_{16-i} for $i = 1, \cdots, 15$. Table 2 is obtained.

Here the NE–SW diagonal is shared by two sets of entries. This is a characteristic feature arising when m is even but not when m is odd.

In fact, going now to the simplest case where $m = 2$ and $f(x) = x^2 - x - 1$, the table which arises is as follows·

	1	2	3
1			0 1 / 2 3
2		0 2 / 3 1	
3	0 3 / 1 2		

For $m = 5$, $\mu = 31$. Take $f(x) = x^5 - x^2 - 1$. Define R_i to be Q_{32-i}. Then we obtain Table 3 (only the first line of entries need be given).

TABLE 3

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
12,20	8,23	9,21	14,15		10,17		27,29				2,19			18,31

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
22,26	5,11						16,25	3,6			13,24		7,28	4,30	0,1

7. If the columns of the first design of Section 5 be regarded as blocks, the rows as a set of treatments $a_1, \cdots, a_7$, and the numbers in the squares as a second set of treatments $b_0, \cdots b_7$, then the design is an incomplete block with respect to the first set of treatments and a complete randomized block with respect to the second set of treatments. The design is also balanced with respect to combinations of different levels of treatment a, with different levels of treatment b. The usual parametric model (Model I) would be

$$x_{tij} = A + B_t + \alpha_i + \beta_j + z_{tij}$$

(where x_{tij} denotes the observation on treatment combination $a_i b_j$ in the tth block, $\sum B_t = \sum \alpha_i = \sum \beta_j = 0$ and the z_{tij}'s are mutually independent random variables with common variance and mean zero). The analysis of variance appropriate to this model is obtained as follows:

(i) Carry out the standard incomplete block analysis on the means $\bar{x}_{ti}$ of pairs of observations for treatments a_i in the same (tth) block. Multiply the resultant sums of squares by two. This will give the Between Blocks and Adjusted Between Treatments a sums of squares in the final table.

(ii) Compute the Between Treatments b sum of squares in the usual way (that is, $7 \sum_{j=0}^{7} (\bar{x}_{..j} - \bar{x}_{...})^2$).

(iii) Compute the Residual sum of squares as Residual in (i) + $\sum_t \sum_i \sum_j (x_{tij} - \bar{x}_{ti.})^2$ − Between Treatment sum of squares in (ii).

The degrees of freedom appropriate to these sums of squares are then

Blocks	6
Adjusted Treatments a	6
Treatments b	7
Residual	36

One advantage of this design lies in the fact that the treatment b sum of squares is orthogonal to the treatment a sum of squares. It is, unfortunately, not possible to test for interaction between the two sets of treatments. Certain specific interactions may, however, be isolated from the Residual sum of squares. For example the contrast b_2 vs. b_4 in the presence of a_1 can be compared with the average effect of the same contrast in the presence of $a_2 a_3 \cdots a_7$, provided it is assumed that other interactions between a and b are negligible. The calculation of the sum of squares for such a contrast could be based on a two-way table with entries

$$b_2 a_1, \quad b_4 a_1 \qquad b_2 \sum_{i=2}^{7} a_i, \qquad b_4 \sum_{i=2}^{7} a_i$$

in the usual way.

Alternatively, the design may be regarded as an incomplete block design for treatments a, with main plots split for treatment b. In this case the design should be regarded as an incomplete block design also with respect to treatments b. The

model becomes

$$x_{tij} = A + B_t + \alpha_i + \beta_j + u_{ti} + z_{tij}$$

where the u_{ti}'s are independent random variables, with zero mean and common variance, which are also independent of the z_{tij}'s. The two incomplete block analyses may be carried out separately (except that the Blocks sum of squares in the Treatments b analysis is the Total sum of squares in the Treatments a analysis). The sums of squares in the complete analysis, and their associated degrees of freedom, are

Blocks	6	
Adjusted Treatments a	6	As in the original
Error (i)	15	analysis (i)
Adjusted Treatments b	7	
Error (ii)	21	

As in the earlier analysis it is not possible, in general, to test for interaction between a and b, but certain specific interactions can be isolated from Error (ii).

Similar considerations apply to the second design of Section 7.

The design shown in paragraph 8 is a supplemented incomplete block design (in the sense of [5]) with respect to treatment a. The analysis of the design will, however, be similar to that described above for the designs of Section 7, and in particular the Treatment b sum of squares will again be orthogonal to the adjusted Treatment a sum of squares.

REFERENCES

[1] T. G. Room, "A new type of magic square," *Math. Gazette*, Vol 39 (1955), p. 307.
[2] J. Singer, "A theorem in finite projective geometry and some applications to number theory," *Trans. Amer. Math Soc*, Vol. 43, (1938), pp. 377–385.
[3] C. C. MacDuffee, *Introduction to Abstract Algebra*, John Wiley & Sons.
[4] L. E. Dickson, *Linear Groups*, Teubner, Leipzig (1901).
[5] T. N. Hoblyn, S. C. Pearce and G. H. Freeman, "Some considerations in the design of successive experiments in fruit plantations," *Biometrics*, Vol. 10, (1954), pp. 503–515

A MULTIVARIATE TCHEBYCHEFF INEQUALITY[1]

By Ingram Olkin[2] and John W. Pratt[3]

University of Chicago

0. Abstract. A multivariate Tchebycheff inequality is given, in terms of the covariances of the random variables in question, and it is shown that the inequality is sharp, i.e., the bound given can be achieved. This bound is obtained from the solution of a certain matrix equation and cannot be computed easily in general. Some properties of the solution are given, and the bound is given explicitly for some special cases. A less sharp but easily computed and useful bound is also given.

1. Introduction and outline. Tchebycheff's inequality states that if y is any real random variable with mean 0 and variance σ^2, then

$$(1.1) \qquad P(|y| \geqq k\sigma) \leqq 1/k^2.$$

Berge [1] has generalized this result as follows. If y_1 and y_2 are any real random variables with means 0, variances σ_1^2 and σ_2^2 respectively, and correlation ρ, then

$$(1.2) \qquad P(|y_1| \geqq k\sigma_1 \quad \text{or} \quad |y_2| \geqq k\sigma_2) \leqq \frac{1 + \sqrt{1-\rho^2}}{k^2}.$$

Berge gives an example where the inequality is achieved.

Suppose $y = (y_1, \cdots, y_p)$ is a random vector with mean 0 and nonsingular covariance matrix Σ. We seek an upper bound, depending on Σ and $k_1, \cdots, k_p$, for $P(|y_i| \geqq k_i\sigma_i$ for some $i)$.

The problem can be reduced by letting $x_i = y_i/(k_i\sigma_i)$. Then $x = (x_i, \cdots, x_p)$ has mean 0 and covariance matrix $\Pi = K^{-1}RK^{-1}$, where $R = (\rho_{ij})$ is the correlation matrix of y (and of x), $\Pi_{ij} = \sigma_{ij}/(\sigma_i\sigma_j k_i k_j) = \rho_{ij}/(k_i k_j)$, and K is a diagonal matrix with diagonal elements $k_1, \cdots, k_p$. Furthermore, $|y_i| \geqq k_i\sigma_i$ if and only if $|x_i| \geqq 1$, so $P(|y_i| \geqq k_i\sigma_i$ for some $i) = P(|x_i| \geqq 1$ for some $i)$.

Suppose A is a $p \times p$ matrix such that

$$(1.3) \qquad xAx' \geqq 1 \quad \text{if} \quad |x_i| = 1 \quad \text{for some } i.$$

Then, looking at scalar multiples of x, we see that

$$(1.4) \qquad xAx' \geqq 1 \quad \text{if} \quad |x_i| \geqq 1 \quad \text{for some } i,$$

Received February 18, 1957; revised June 11, 1957.

[1] This research was sponsored by the Office of Ordnance Research, U.S. Army, and the Statistics Branch, Office of Naval Research.

[2] Work done while on leave from Michigan State University.

[3] Now at Harvard University.

and that

$$(1.5) \qquad xAx' \geqq 0 \quad \text{for all } x,$$

i.e., A is positive definite. Therefore

LEMMA 1.1. *If A satisfies* (1.3), *then*

$$(1.6) \quad P(|y_i| \geqq k_i\sigma_i \text{ for some } i) = P(|x_i| \geqq 1 \text{ for some } i) \leqq E(xAx') = \operatorname{tr} A\Pi,$$

where tr *denotes trace.*

Each A satisfying (1.3) therefore gives an upper bound for

$$P(|x_i| \geqq 1 \text{ for some } i).$$

The smallest bound obtainable in this way is the minimum of tr $A\Pi$ over all A satisfying (1.3). The set $\mathfrak{A}$ of all such matrices A is obviously convex, closed, and bounded from below, and tr $A\Pi$ is linear in A, so this minimum is achieved at an extreme point of $\mathfrak{A}$. In Theorem 3.3 it is shown that A is an extreme point of $\mathfrak{A}$ if and only if A^{-1} is positive definite and has 1's *on the main diagonal*. Furthermore, there is a unique extreme point of $\mathfrak{A}$ minimizing tr $A\Pi$, namely that extreme point A such that $A\Pi A$ is diagonal (Theorem 3.5). The *bound thus obtained is the best possible, inasmuch as, if it is less than* 1, *there* is a distribution for x (with mean 0 and covariance matrix Π) under which it is achieved, and otherwise there is a distribution for x under which

$$P(|x_i| \geqq 1 \text{ for some } i) = 1$$

(Theorem 3.7).

The minimizing matrix is easy to compute explicitly only in some special cases (Sec. 5). In the case $p = 2$, $k_1 = k_2 = k$, Berge lets $A = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}^{-1}$, shows that A satisfies (1.3), and minimizes tr $A\Pi$ with respect to a. Following this lead, in Sec. 2 we let $A = [(1 - a)I + ae'e]^{-1}$, where $e = (1, \cdots, 1)$, show that A satisfies (1.3) for $1 > a > -1/(p - 1)$, *and minimize* tr $A\Pi$ *with re*spect to a, obtaining the bound in Theorem 2.3. Though the minimum over such A is in general, except in the case $p = 2$, not the minimum over all A satisfying (1.3), it provides a useful and easily computed bound. Lal [3] considers a matrix similar in form to that of Sec. 2. However, this does not lead to the best bound, as Lal asserts, and indeed his bound is not as tight as that given in Theorem 2.3 unless $p = 2$ or $\rho_{ij} = 0$ for all $i \neq j$.

2. A multivariate inequality. We will now carry out the program of the last paragraph.

LEMMA 2.1. $A = [(1 - a)I + ae'e]^{-1}$ *satisfies* (1.3) *if* $1 > a > -1/(p - 1)$.

PROOF. $A = [(1-a)I + ae'e]^{-1} = (I - \alpha e'e)/(1-a)$, *where* $\alpha = a/[1+(p-1)a]$. $x[I - \alpha e'e]x' = \sum x_i^2 - \alpha(\sum x_i)^2$

$$\geqq \begin{cases} \sum x_i^2 & \text{if} \quad 0 \geqq a \geqq -1/(p-1), \quad \text{i.e.,} \quad \alpha \leqq 0; \\ (1-p\alpha)\sum x_i^2 & \text{if} \quad 0 \leqq a < 1, \quad \text{i.e.,} \quad 0 \leqq \alpha < 1/p \end{cases}$$

(The second case follows from $(\sum x_i)^2 \leqq p\sum x_i^2$.) The right-hand side becomes infinite with $\sum x_i^2$, so the minimum over all $(p-1)$ – vectors z of

$$(1, z)(I - \alpha e'e)(1, z)'$$

occurs at a finite z. Differentiating

$$(1, z)(I - \alpha e'e)(1, z)' = 1 + \sum z_i^2 - \alpha(1 + \sum z_i)^2$$

with respect to each z_i we find that the minimizing z must satisfy $2z_i - 2\alpha(1 + \sum z_j) = 0$ for all i, or $z - \alpha z e'e - \alpha e = 0$. (Here e has $p-1$ coordinates.) It follows that all z_i are equal, and that $\sum z_i = (p-1)a$, so $z = ae$. Therefore the minimum over z of $(1, z)(I - \alpha e'e)(1, z)'$ is $1 - a$, and thus the minimum over z of

$$(1, z)A(1, z)'$$

is 1. The lemma follows. (See also Lemma 5.1.) || (This symbol will be used to indicate the end of a proof.)

LEMMA 2.2. tr $[(1-a)I + ae'e]^{-1}\Pi$ *is minimized for* $1 > a > -1/(p-1)$ *by*

$$a = \frac{t - \sqrt{u(pt-u)/(p-1)}}{u-(p-1)t}, \tag{2.1}$$

where $t = \operatorname{tr}\Pi = \sum \Pi_{ii} = \sum 1/k_i^2$ *and* $u = e\Pi e' = \sum \Pi_{ij} = \sum \rho_{ij}/(k_i k_j)$.

PROOF. tr $[(1-a)I + ae'e]^{-1}\Pi = \operatorname{tr}(I - \alpha e'e)\Pi/(1-a) = (t - \alpha u)/(1-a)$. The derivative of this quantity with respect to a has zeros at

$$a = \frac{t \pm \sqrt{u(pt-u)/(p-1)}}{u-(p-1)t}.$$

The condition $1 > a > -1/(p-1)$ is satisfied if and only if

$$\mp\sqrt{u(pt-u)/(p-1)}$$

is between $u/(p-1)$ and $(pt-u)$. The upper sign is impossible because

$$u/(p-1) \quad \text{and} \quad (pt-u)$$

are both positive. The lower sign is possible because $\sqrt{u(pt-u)/(p-1)}$ is the geometric mean of $u/(p-1)$ and $(pt-u)$. The extremum is a minimum since $(t-\alpha u)/(1-a) \to \infty$ as $a \to 1$ or $a \to -1/(p-1)$. ||

Substituting (2.1) in (1.6) and simplifying, we obtain, by Lemmas 1.1, 2.1, and 2.2,

THEOREM 2.3. *$P(|y_i| \geqq k_i\sigma_i$ for some $i) = P(|x_i| \geqq 1$ for some $i)$*

$$\leqq \frac{p-1}{p} t - \frac{p-2}{p^2} u + \frac{2}{p^2} \sqrt{u(pt-u)(p-1)}$$
$$= [\sqrt{u} + \sqrt{(pt-u)(p-1)}]^2/p^2.$$

In the case $p = 2$, we obtain

$$P(|y_1| \geqq k_1\sigma_1 \text{ or } |y_2| \geqq k_2\sigma_2) \leqq \frac{1}{2k_1^2k_2^2} [k_1^2 + k_2^2 + \sqrt{(k_1^2 + k_2^2)^2 - 4\rho^2 k_1^2 k_2^2}],$$

which is Lal's equation (B), and is to be compared with Berge's result, (1.2).

3. The sharpest inequality. In this section we seek the tightest bound obtainable from Lemma 1.1, and show that it is sharp, following the outline in the next-to-last paragraph of Sec. 1. What we seek, then, is the minimum of tr $A\Pi$ for A satisfying (1.3), i.e., for $A \varepsilon \mathfrak{A}$. As remarked before, the minimum occurs at an extreme point of $\mathfrak{A}$. We start by characterizing, in Lemma 3.2, the matrices in $\mathfrak{A}$, and, in Theorem 3 3, the extreme points of $\mathfrak{A}$ We use the following lemma, which has some independent interest.

LEMMA 3.1. *If A is positive definite, the minimum of xAx' for $x_1 = 1$ is $1/b_{11}$ and occurs at $(1, b/b_{11})$, and only there, where*

$$B = \begin{pmatrix} b_{11} & b \\ b' & B_{22} \end{pmatrix} = A^{-1} = \begin{pmatrix} a_{11} & a \\ a' & A_{22} \end{pmatrix}^{-1}.$$

PROOF. It is easily checked that

$$b_{11} = (a_{11} - aA_{22}^{-1}a')^{-1}, \qquad b = -b_{11}aA_{22}^{-1}, \qquad B_{22} = A_{22}^{-1} + A_{22}^{-1}a'b_{11}aA_{22}^{-1}.$$

"Completing the square," we have

$$(1, z)A(1, z)' = a_{11} + 2az' + zA_{22}z'$$
$$= a_{11} - aA_{22}^{-1}a' + (z + aA_{22}^{-1})A_{22}(z + aA_{22}^{-1})'$$
$$= b_{11}^{-1} + (z - b_{11}^{-1}b)A_{22}(z - b_{11}^{-1}b)'.$$

Since A_{22} is positive definite, the lemma follows. Alternatively, $(1, z)A(1, z)'$ could be differentiated with respect to each coordinate of z, as in the proof of Lemma 2.1. ||

It follows from this lemma and (1.5) that

LEMMA 3.2. *$A \varepsilon \mathfrak{A}$ if and only if $B = A^{-1}$ is positive definite and $b_{ii} \leqq 1$, $i = 1, \cdots, p$.*

THEOREM 3.3. *A is extreme in $\mathfrak{A}$ if and only if $B = A^{-1}$ is positive definite and $b_{ii} = 1$, $i = 1, \cdots, p$.*

PROOF. (i) Suppose B is positive definite and all $b_{ii} = 1$. Then, by Lemma 3.2, $A \varepsilon \mathfrak{A}$. Suppose $A = (A_1 + A_2)/2$, $A_1 \varepsilon \mathfrak{A}$, $A_2 \varepsilon \mathfrak{A}$. For each i, by Lemma

PROOF. $A = [(1 - a)I + ae'e]^{-1} = (I - \alpha e'e)/(1 - a)$, *where* $\alpha = a/[1 + (p - 1)a]$. $x[I - \alpha e'e]x' = \sum x_i^2 - \alpha(\sum x_i)^2$

$$\geqq \begin{cases} \sum x_i^2 & \text{if } 0 \geqq a \geqq -1/(p-1), \quad \text{i.e.,} \quad \alpha \leqq 0; \\ (1 - p\alpha)\sum x_i^2 & \text{if } 0 \leqq a < 1, \quad \text{i.e.,} \quad 0 \leqq \alpha < 1/p \end{cases}$$

(The second case follows from $(\sum x_i)^2 \leqq p\sum x_i^2$.) The right-hand side becomes infinite with $\sum x_i^2$, so the minimum over all $(p - 1)$ − vectors z of

$$(1, z)(I - \alpha e'e)(1, z)'$$

occurs at a finite z. Differentiating

$$(1, z)(I - \alpha e'e)(1, z)' = 1 + \sum z_i^2 - \alpha(1 + \sum z_i)^2$$

with respect to each z_i we find that the minimizing z must satisfy $2z_i - 2\alpha(1 + \sum z_j) = 0$ for all i, or $z - \alpha z e'e - \alpha e = 0$. (Here e has $p - 1$ coordinates.) It follows that all z_i are equal, and that $\sum z_i = (p - 1)a$, so $z = ae$. Therefore the minimum over z of $(1, z)(I - \alpha e'e)(1, z)'$ is $1 - a$, and thus the minimum over z of

$$(1, z)A(1, z)'$$

is 1. The lemma follows. (See also Lemma 5.1.) || (This symbol will be used to indicate the end of a proof.)

LEMMA 2.2. tr $[(1 - a)I + ae'e]^{-1}\Pi$ *is minimized for* $1 > a > -1/(p - 1)$ *by*

$$a = \frac{t - \sqrt{u(pt - u)/(p - 1)}}{u - (p - 1)t}, \tag{2.1}$$

where $t = \text{tr}\,\Pi = \sum \Pi_{ii} = \sum 1/k_i^2$ *and* $u = e\Pi e' = \sum \Pi_{ij} = \sum \rho_{ij}/(k_i k_j)$.

PROOF. $\text{tr}\,[(1 - a)I + ae'e]^{-1}\Pi = \text{tr}\,(I - \alpha e'e)\Pi/(1 - a) = (t - \alpha u)/(1 - a)$. The derivative of this quantity with respect to a has zeros at

$$a = \frac{t \pm \sqrt{u(pt - u)/(p - 1)}}{u - (p - 1)t}.$$

The condition $1 > a > -1/(p - 1)$ is satisfied if and only if

$$\mp\sqrt{u(pt - u)/(p - 1)}$$

is between $u/(p - 1)$ and $(pt - u)$. The upper sign is impossible because

$$u/(p - 1) \qquad \text{and} \qquad (pt - u)$$

are both positive. The lower sign is possible because $\sqrt{u(pt - u)/(p - 1)}$ is the geometric mean of $u/(p - 1)$ and $(pt - u)$. The extremum is a minimum since $(t - \alpha u)/(1 - a) \to \infty$ as $a \to 1$ or $a \to -1/(p - 1)$. ||

Substituting (2.1) in (1.6) and simplifying, we obtain, by Lemmas 1.1, 2.1, and 2.2,

THEOREM 2.3. *$P(|y_i| \geqq k_i\sigma_i$ for some $i) = P(|x_i| \geqq 1$ for some $i)$*

$$\leqq \frac{p-1}{p} t - \frac{p-2}{p^2} u + \frac{2}{p^2} \sqrt{u(pt-u)(p-1)}$$
$$= [\sqrt{u} + \sqrt{(pt-u)(p-1)}]^2/p^2.$$

In the case $p = 2$, we obtain

$$P(|y_1| \geqq k_1\sigma_1 \text{ or } |y_2| \geqq k_2\sigma_2) \leqq \frac{1}{2k_1^2k_2^2}[k_1^2 + k_2^2 + \sqrt{(k_1^2+k_2^2)^2 - 4\rho^2k_1^2k_2^2}],$$

which is Lal's equation (B), and is to be compared with Berge's result, (1 2).

3. The sharpest inequality. In this section we seek the tightest bound obtainable from Lemma 1.1, and show that it is sharp, following the outline in the next-to-last paragraph of Sec. 1. What we seek, then, is the minimum of tr $A\Pi$ for A satisfying (1.3), i.e., for $A \varepsilon \mathfrak{A}$. As remarked before, the minimum occurs at an extreme point of $\mathfrak{A}$. We start by characterizing, in Lemma 3.2, the matrices in $\mathfrak{A}$, and, in Theorem 3.3, the extreme points of $\mathfrak{A}$. We use the following lemma, which has some independent interest

LEMMA 3.1. *If A is positive definite, the minimum of xAx' for $x_1 = 1$ is $1/b_{11}$ and occurs at $(1, b/b_{11})$, and only there, where*

$$B = \begin{pmatrix} b_{11} & b \\ b' & B_{22} \end{pmatrix} = A^{-1} = \begin{pmatrix} a_{11} & a \\ a' & A_{22} \end{pmatrix}^{-1}.$$

PROOF. It is easily checked that

$$b_{11} = (a_{11} - aA_{22}^{-1}a')^{-1}, \qquad b = -b_{11}aA_{22}^{-1}, \qquad B_{22} = A_{22}^{-1} + A_{22}^{-1}a'b_{11}aA_{22}^{-1}.$$

"Completing the square," we have

$$(1, z)A(1, z)' = a_{11} + 2az' + zA_{22}z'$$
$$= a_{11} - aA_{22}^{-1}a' + (z + aA_{22}^{-1})A_{22}(z + aA_{22}^{-1})'$$
$$= b_{11}^{-1} + (z - b_{11}^{-1}b)A_{22}(z - b_{11}^{-1}b)'.$$

Since A_{22} is positive definite, the lemma follows. Alternatively, $(1, z)A(1, z)'$ could be differentiated with respect to each coordinate of z, as in the proof of Lemma 2.1. ||

It follows from this lemma and (1.5) that

LEMMA 3.2. *$A \varepsilon \mathfrak{A}$ if and only if $B = A^{-1}$ is positive definite and $b_{ii} \leqq 1$, $i = 1, \cdots, p$.*

THEOREM 3.3. *A is extreme in $\mathfrak{A}$ if and only if $B = A^{-1}$ is positive definite and $b_{ii} = 1, i = 1, \cdots, p$.*

PROOF. (i) Suppose B is positive definite and all $b_{ii} = 1$. Then, by Lemma 3.2, $A \varepsilon \mathfrak{A}$. Suppose $A = (A_1 + A_2)/2$, $A_1 \varepsilon \mathfrak{A}$, $A_2 \varepsilon \mathfrak{A}$. For each i, by Lemma

3.1,

$$1 = 1/b_{ii} = \min_{x_i=1} xAx' \geqq \tfrac{1}{2}\,[\min_{x_i=1} xA_1\,x' + \min_{x_i=1} xA_2\,x'],$$

$$\min_{x_i=1} xA_1\,x' \geqq 1. \qquad \min_{x_i=1} xA_2\,x' \geqq 1.$$

It follows that

$$\min_{x_i=1} xA_1\,x' = 1 = \min_{x_i=1} xA_2\,x',$$

and the minima occur at the same point. This implies, by Lemma 3.1, that the ith row of A_1^{-1} equals the ith row of A_2^{-1}. As this is true for each i, $A_1 = A_2$. Therefore A is extreme in $\mathfrak{A}$, which proves the "if".

(ii) If B is not positive definite, $A \; \varepsilon \; \mathfrak{A}$, by Lemma 3.2. Suppose B is positive definite but $b_{ii} < 1$ for some i, say $b_{11} < 1$. Let

$$B(\delta) = B + \begin{pmatrix} \delta & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} b_{11} + \delta & b \\ b' & B_{22} \end{pmatrix}.$$

By Lemma 3.2, $B^{-1}(\delta) \; \varepsilon \; \mathfrak{A}$ for δ small enough. If we can choose $\delta_1 \neq \delta_2$ such that $B^{-1}(\delta_1) \; \varepsilon \; \mathfrak{A}$, $\quad B^{-1}(\delta_2) \; \varepsilon \; \mathfrak{A}$, and

$$A = B^{-1} = \theta B^{-1}(\delta_1) + (1 - \theta)B^{-1}(\delta_2) \tag{3.1}$$

for some θ, $0 < \theta < 1$, we will have shown that A is not extreme in $\mathfrak{A}$.

According to the first sentence of the proof of Lemma 3.1, with A and B interchanged, $B^{-1}(\delta)$ is a linear function of its upper left element $a_{11}(\delta)$, so (3.1) is equivalent to

$$a_{11} = a_{11}(0) = \theta a_{11}(\delta_1) + (1 - \theta)a_{11}(\delta_2).$$

Furthermore,

$$a_{11}\,(\delta) = \frac{1}{b_{11} + \delta - bB_{22}^{-1}\,b'} = \frac{1}{\delta + 1/a_{11}} = \frac{a_{11}}{1 + \delta a_{11}}.$$

Therefore (3.1) is equivalent to

$$\frac{\theta\delta_1}{1 + \delta_1\,a_{11}} + \frac{(1 - \theta)\delta_2}{1 + \delta_2\,a_{11}} = 0,$$

and it is clear that δ_1 and δ_2 can be chosen as desired. ||

This reduces the problem to that of minimizing $\operatorname{tr} B^{-1}\Pi$ for $B \; \varepsilon \; \mathfrak{B}$, where $\mathfrak{B}$ is the set of positive definite matrices with ones on the main diagonal. We will now show that $\operatorname{tr} B^{-1}\Pi$ is minimized at a unique interior point $\bar{B}$ of $\mathfrak{B}$, (Theorem 3.4), and characterize $\bar{B}$ (Theorem 3.5).

THEOREM 3.4. *$\operatorname{tr} B^{-1}\Pi$ is a strictly convex function of B for $B \; \varepsilon \; \mathfrak{B}$, and has a unique minimum, which occurs at an interior point $\bar{B}$ of $\mathfrak{B}$.*

PROOF. Let $B(t)$ be a straight line in $\mathfrak{B}$. Then dB/dt is a symmetric matrix,

$d^2B/dt^2 = 0$, and

$$\frac{d}{dt} \operatorname{tr} B^{-1}\Pi = -\operatorname{tr} B^{-1}\left(\frac{dB}{dt}\right) B^{-1}\Pi,$$

$$\frac{d^2}{dt^2} \operatorname{tr} B^{-1}\Pi = 2 \operatorname{tr} B^{-1}\left(\frac{dB}{dt}\right) B^{-1}\left(\frac{dB}{dt}\right) B^{-1}\Pi > 0.$$

This proves the strict convexity. The rest follows, since $\mathfrak{B}$ is convex and bounded, and $\operatorname{tr} B^{-1}\Pi \to \infty$ as B approaches the boundary of $\mathfrak{B}$. The latter follows from the fact that

$$\operatorname{tr} B^{-1}\Pi \geqq (\operatorname{tr} B^{-1})(\text{smallest eigenvalue of } \Pi). \|$$

THEOREM 3.5. *$\bar{B}$ is the unique point of $\mathfrak{B}$ such that $\bar{B}^{-1}\Pi\bar{B}^{-1}$, or equivalently $\bar{B}\Pi^{-1}\bar{B}$, is diagonal.*

PROOF. By Theorem 3.4, $\bar{B}$ is the unique point of $\mathfrak{B}$ for which

$$\frac{d}{db_{ij}} \operatorname{tr} B^{-1}\Pi = \operatorname{tr} B^{-1}\left(\frac{dB}{db_{ij}}\right) B^{-1}\Pi = \operatorname{tr}\left(\frac{dB}{db_{ij}}\right) B^{-1}\Pi B^{-1} = 2c_{ij} = 0$$

for $i \neq j$, where $C = B^{-1}\Pi B^{-1}$, and dB/db_{ij} is a matrix with all elements zero except the (i, j)-th and (j, i)-th, which are one ||

We note that $B^{-1}\Pi B^{-1} = C$ if and only if

$$B = \Pi^{1/2}(\Pi^{1/2}C\Pi^{1/2})^{-1/2}\Pi^{1/2} = C^{-1/2}(C^{1/2}\Pi C^{1/2})^{1/2}C^{-1/2}.$$

By Theorems 3.3, 3.4, and 3.5, the tightest inequality obtainable from Lemma 1.1 is

THEOREM 3.6. $P(|y_i| \geqq k_i\sigma_i$ *for some* $i) = P(|x_i| \geqq 1$ *for some* $i)$

$$\leqq \operatorname{tr} \bar{B}^{-1}\Pi = \operatorname{tr} \bar{B}^{-1}\Pi\bar{B}^{-1},$$

where $\bar{B}$ is the unique positive definite matrix having ones on the main diagonal such that $\bar{B}\Pi^{-1}\bar{B}$ is diagonal.

We note that $\operatorname{tr} \bar{B}^{-1}\Pi = \operatorname{tr}(\bar{B}^{-1}\Pi\bar{B}^{-1})\bar{B} = \operatorname{tr} \bar{B}^{-1}\Pi\bar{B}^{-1}$, since $\bar{B}^{-1}\Pi\bar{B}^{-1}$ is diagonal and $\bar{B}$ has ones on the diagonal.

According to the following theorem, the bound given in Theorem 3.6 is the smallest possible bound except when the smallest possible bound is the trivial bound 1.

THEOREM 3.7. *Let $\Theta = \bar{B}^{-1}\Pi\bar{B}^{-1}$ and $\theta_1, \cdots, \theta_p$ be its diagonal elements. Then*

$$\operatorname{tr} \bar{B}^{-1}\Pi = \operatorname{tr} \bar{B}^{-1}\Pi\bar{B}^{-1} = \operatorname{tr} \Theta = \sum \theta_i .$$

If $\sum \theta_i \leqq 1$, equality holds in Theorem 3.6 if and only if

$$P(x = b^i) = P(x = -b^i) = \theta_i/2, \qquad i = 1, \cdots, p,$$

(3.2)

$$P(x = 0) = 1 - \sum \theta_i ,$$

where $b^1, \cdots, b^p$ *are the rows of* $\bar{B}$. *If* $\sum \theta_i > 1$, $P(|x_i| \geqq 1$ *for some* $i) = 1$ *if*

$$P(x = \sqrt{\textstyle\sum \theta_i}\, b^i) = P(x = -\sqrt{\textstyle\sum \theta_i}\, b^i) = \theta_i/(2 \textstyle\sum \theta_i), \quad i = 1, \cdots, p. \tag{3.3}$$

Proof. If $\sum \theta_i \leqq 1$, (3.2) is a distribution for x, and if x has this distribution, equality holds in Theorem 3.6. If x has the distribution (3.3) and $\sum \theta_i > 1$, then, with probability one, $|x_i| \geqq \sqrt{\sum \theta_i} > 1$ for some i. In either case, x has mean 0 and covariance matrix

$$E(x'x) = \sum \theta_i b^{i\prime} b^i = \bar{B}\Theta\bar{B} = \Pi.$$

This proves the "if".

It remains to prove the "only if". Suppose $\sum \theta_i \leqq 1$ and equality holds in Theorem 3.6. Then, by the relation of (1.6) to (1.4) and (1.5), with probability one,

$$x\bar{B}^{-1}x' = 1 \quad \text{if} \quad |x_i| \geqq 1 \quad \text{for some } i,$$

and

$$x\bar{B}^{-1}x' = 0 \quad \text{otherwise}.$$

It follows, by Lemma 3.1, that the distribution of x is concentrated at 0 and $\pm b^1, \cdots, \pm b^p$. Then

$$E(x) = \sum [P(x = b^i) - P(x = -b^i)]b^i.$$

But $E(x) = 0$ and $b^1, \cdots, b^p$ are linearly independent, since they are the rows of a non-singular matrix, so $P(x = b^i) = P(x = -b^i)$ for all i. Then

$$E(x'x) = \sum 2P(x = b^i) b^{i\prime} b^i = \bar{B}D\bar{B},$$

where D is a diagonal matrix with diagonal elements

$$2P(x = b^1), \cdots, 2P(x = b^p).$$

But

$$E(x'x) = \Pi, \quad \text{so} \quad D = \bar{B}^{-1}\Pi\bar{B}^{-1} = \Theta,$$

and (3.2) follows. ||

4. On the solution of $\bar{B}\Theta\bar{B} = \Pi$. From $\Pi = \bar{B}\Theta\bar{B}$, we find that

$$\Pi_{ij} = \textstyle\sum_\alpha \bar{b}_{i\alpha}\theta_\alpha \bar{b}_{\alpha j},$$

and for $i = j$ we have the system of equations

$$1/k_i^2 = \sum_\alpha \bar{b}_{i\alpha}^2 \theta_\alpha, \qquad i = 1, \cdots, p.$$

If we write $\bar{B} \times \bar{B} = (\bar{b}_{ij}^2)$, then

$$(\theta_1, \cdots, \theta_p) = (k_1^{-2}, \cdots, k_p^{-2})(\bar{B} \times \bar{B})^{-1}.$$

Thus given $\bar{B}$ and $k_1, \cdots, k_p$, we can solve for Θ and Π. The matrix $B \times B$ is the Hadamard product, and is positive definite if B is ([2], p. 143). Given $k_1, \cdots, k_p$, $\bar{B}$ results from some Π if and only if $\bar{B} \varepsilon \mathcal{B}$ and

$$(k_1^{-2}, \cdots, k_p^{-2})(\bar{B} \times \bar{B})^{-1}$$

has positive elements. The following example shows that this last condition is not automatically satisfied.

$$B = \begin{pmatrix} 1 & .8 & .8 \\ .8 & 1 & .5 \\ .8 & .5 & 1 \end{pmatrix}, \qquad |B \times B|(B \times B)^{-1} = \begin{pmatrix} .9375 & -.4800 & -.4800 \\ -.4800 & .5904 & .1596 \\ -.4800 & .1596 & .5904 \end{pmatrix},$$

$$k_1 = \cdots = k_p = 1.$$

Every $\bar{B} \varepsilon \mathcal{B}$ results from some $k_1, \cdots, k_p$ and Π, e.g., for

$$(k_1^{-2}, \cdots, k_p^{-2}) = (1, \cdots, 1)\bar{B} \times \bar{B}$$

This section began with a procedure for determining Π from $\bar{B}$ by standard matrix operations. It appears that $\bar{B}$ cannot be obtained from Π by standard matrix operations except in special cases. We now give two properties of the solution (Theorems 4.1 and 4.2)

THEOREM 4.1. *If P is a permutation matrix and $P\Pi P = \Pi$, then $P\bar{B}P = \bar{B}$.*

PROOF.

$$(P\bar{B}P)\Pi^{-1}(P\bar{B}P) = P\bar{B}\Pi^{-1}\bar{B}P = P\Theta^{-1}P = \Theta^{-1}$$

$P\bar{B}P \varepsilon \mathcal{B}$, so by the uniqueness in Theorem 3.5, $P\bar{B}P = \bar{B}$. ||

THEOREM 4.2. *If* $\Pi = \begin{pmatrix} \Pi_1 & 0 \\ 0 & \Pi_2 \end{pmatrix}$, *then* $\bar{B} = \begin{pmatrix} \bar{B}_1 & 0 \\ 0 & \bar{B}_2 \end{pmatrix}$, *where $\bar{B}_i$ minimizes*

$$\operatorname{tr} \bar{B}_i \Pi_i^{-1} \quad \text{in} \quad \mathcal{B}_i, \qquad i = 1, 2.$$

PROOF. If $\bar{B}_i\Pi_i^{-1}\bar{B}_i$ is diagonal, $i = 1, 2$, then $\bar{B}\Pi^{-1}\bar{B}$ is diagonal, and by the uniqueness of $\bar{B}$, the conclusion follows ||

5. Special cases.

THEOREM 5.1. *If $\Pi^{1/2}$ has equal diagonal elements, say, d, then*

$$\bar{B} = \Pi^{1/2}/d, \qquad \Theta = d^2 I$$

and

$$P(|y_i| \geqq k_i\sigma_i \text{ for some } i) = P(|x_i| \geqq 1 \text{ for some } i) \leqq \operatorname{tr} \bar{B}^{-1}\Pi = d^2 p.$$

This follows from Theorem 3.5. (The result for singular Π is an easy consequence of the result for non-singular Π.)

We note that $\Pi^{1/2}$ has equal diagonal elements if the group of permutation matrices P such that $P\Pi P = \Pi$ is transitive, i.e., every coordinate of x can be carried into every other one by a permutation of coordinates which preserves

ne covariances, i.e., $k_1 = \cdots = k_p$, and every coordinate of y can be carried nto every other by a permutation of coordinates which preserves the correla-ons. This follows from the fact that $P\Pi^{1/2}P = \Pi^{1/2}$ if $P\Pi P = P$, since then $(P\Pi^{1/2}P)^2 = P\Pi P = \Pi$.

$\bar{B} = (1 - a)I + ac'c$, i.e., the inequality of Sec. 2 is the best possible, if and nly if the elements of Π are

$$\Pi_{ii} = 1/k_i^2,$$

5.1)
$$\Pi_{ij} = \rho_{ij}/k_i k_j = \frac{a}{1+a}\left[k_i^{-2} + k_j^{-2} + \frac{a(1-a)}{1+(p-1)a^2}\sum k_\alpha^{-2}\right],$$

n which case

5.2)
$$P(|y_i| \geqq k_i\sigma_i \text{ for some } i) \leqq \operatorname{tr} \bar{B}^{-1}\Pi = \sum k_i^{-2}/[1 + (p-1)a^2].$$

In the case $p = 2$, Π is always of this form and (5.2) yields (2.6).

If $k_1 = \cdots = k_p = k$, and $\Pi_{ii} = 1/k^2$, $\Pi_{ij} = \rho/k^2$, then Π is of the form 5.1) and

$$P(|y_i| \geqq k\sigma_i \text{ for some } i) \leqq \operatorname{tr} \bar{B}^{-1}\Pi$$

5.3)
$$= \frac{p}{k^2[1 + (p-1)a^2]} = \frac{[(p-1)\sqrt{1-\rho} + \sqrt{1+(p-1)\rho}]^2}{pk^2}.$$

This could also be obtained from Theorem 2.3, or from Theorem 5.1.

$$\Pi^{1/2} = \frac{\sqrt{1-\rho}}{k} I + \frac{[\sqrt{1+p(-1)p} - \sqrt{1-\rho}]}{kp} e'c.$$

For special values of p and ρ we obtain in addition to Berge's result (1.2), he following inequalities.

(i) $\rho = 1$: $P(|y_i| \geqq k\sigma_i$ for some $i) \leqq 1/k^2$, which amounts to the univariate Tchebycheff inequality.

(ii) $\rho = 0$: For p *uncorrelated* random variables,

$$P(|y_i| \geqq k_i\sigma_i \text{ for some } i) \leqq \sum k_i^{-2},$$

whereas for p *independent* random variables, the univariate Tchebycheff in-equality yields the bound $1 - \prod_{i=1}^{p}(1 - k_i^{-2})$.

(iii) $\rho = -1/(p-1)$: $P(|y_i| \leqq k\sigma_i$ for some $i) \leqq (p-1)/k^2$.

Acknowledgements. We wish to acknowledge the very helpful comments and suggestions of W. Hoeffding and L. J. Savage.

REFERENCES

[1] P. O. Berge, "A note on a form of Tchebycheff's theorem for two variables," *Biometrika*, Vol. 29 (1937), pp. 405–406.

[2] P. R. Halmos, "Finite dimensional vector spaces," *Annals of Mathematics Studies*, No. 7, Princeton University Press, Princeton, 1947.

[3] D. N. Lal, "A note on a form of Tchebycheff's inequality for two or more variables," *Sankhyā*, Vol. 15 (1955), pp. 317–320.

ON SELECTING A SUBSET WHICH CONTAINS ALL POPULATIONS BETTER THAN A STANDARD

By Shanti S. Gupta[1] and Milton Sobel

Bell Telephone Laboratories

1. Summary. A procedure is given for selecting a subset such that the probability that all the populations better than the standard are included in the subset is equal to or greater than a predetermined number P^*. Section 3 deals with the problem of the location parameter for the normal distribution with known and unknown variance. Section 4 deals with the scale parameter problem for the normal distribution with known and unknown mean as well as the chi-square distribution. Section 5 deals with binomial distributions where the parameter of interest is the probability of failure on a single trial In each of the above cases the case of known standard and unknown standard are treated separately. Tables are available for some problems; in other problems transformations are used such that the given tables are again appropriate.

2. Introduction. C. W. Dunnett [3] has considered a different but related problem of comparing several treatment means with a control mean for normal distributions with a common unknown variance His goal is to separate those treatments which are better than the control from those that are worse (or not better). He controls the probability of selecting the standard as the best (i e., classifying all other treatments as worse) *when the treatments are all equal to (or worse than) the standard.* Earlier, E. Paulson [8] considered the problem of selecting the best one of k categories when comparing $k - 1$ categories with a standard. He deals with means of normal distributions with a common unknown variance and also with binomial distributions. He controls the probability of selecting the standard as the best *when the categories are equal to (or worse than) the standard.*

The procedure described in this paper controls the probability that the selected subset *contains all* those populations better than the control *for any possible true configuration.* If we define a correct decision as a selected subset which *contains all* those populations better than the standard, then the procedure given below guarantees a probability of a correct decision to be at least P^*, not only when the $k - 1$ populations are equal to (or worse than) the standard, but for any possible true configuration. Although we are comparing the procedure with the work noted above, it should be stressed that the goals are different and the procedures are not interchangeable. It should be noted that the treatment of Secs. 3 and 4 could be applied to several other distributions in the Koopman-Darmois family.

The goal treated in this paper is more flexible in that it allows the experimenter to choose a subset and withhold judgment about which is *the best one.* Then, if

Received March 14, 1957; revised July 1, 1957.

[1] Now at the University of Alberta

he best one is desired it can be chosen from the selected subset on the basis of economic or other considerations.

Although the title and discussion above use the phraseology "populations better than a standard" we shall actually be interested in selecting all populations as good as or better than the standard; for practical purposes the distinction is of minor importance since in most of the practical problems the parameters of interest can have any value in some interval and are very rarely equal.

To discuss confidence statements we consider first the problems below in which the better populations are the ones with the larger values of the main parameter of interest τ. After the experiment is performed, we can make with confidence P^* the joint statement that *for all populations which are eliminated the parameter value is less than that of the standard*. This joint confidence statement follows from the fact that in selecting a subset *containing all* populations as good as or better than a standard we are automatically eliminating a subset containing only populations worse than the standard. Hence this procedure can be used *to eliminate* those populations which are distinctly inferior to the standard.

For the case in which the better populations are defined to be the ones with the smaller values of τ, the statistical problem is identical and all the results and tables of this paper apply with the obvious modifications.

3. Location parameter—normal populations. We shall assume that populations $\Pi_1, \Pi_2, \cdots, \Pi_p$ with unknown means $\mu_1, \mu_2, \cdots, \mu_p$, respectively are given and that Π_0 is the standard or control, whose mean μ_0 may or may not be known. For clarity we shall discuss the various cases separately.

Case A. Common known variance (μ_0 known). From each of the p populations $\Pi_i (i = 1, 2, \cdots, p)$, n_i independent observations are taken. Let $\bar{x}_i$ denote the sample mean from Π_i and let σ^2 be the common known variance.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i (i = 1, 2, \cdots, p)$ for which

$$\bar{x}_i \geqq \mu_0 - d\sigma/\sqrt{n_i}\text{."} \tag{3.1}$$

To determine the value of d let p_1, p_2 denote the true number of populations with $\mu \geqq \mu_0$ and $\mu < \mu_0$, respectively, so that $p_1 + p_2 = p$. Then the probability P of retaining all the p_1 populations with $\mu \geqq \mu_0$ is given by

$$\begin{aligned} P &= \prod_{i=1}^{p_1} P\{\bar{x}'_i \geqq \mu_0 - d\sigma/\sqrt{n'_i}\} \\ &= \prod_{i=1}^{p_1} P\{\sqrt{n'_i}(\bar{x}'_i - \mu'_i)/\sigma \geqq -d + \sqrt{n'_i}(\mu_0 - \mu'_i)/\sigma\}, \end{aligned} \tag{3.2}$$

where primes refer to values associated with the p_1 populations for which $\mu \geqq \mu_0$. Hence

$$P = \prod_{i=1}^{p_1} \{1 - F(-d + \sqrt{n'_i}(\mu_0 - \mu'_i)/\sigma)\} \tag{3.3}$$

where $F(x)$ refers to the standard normal cumulative distribution function. The μ_i' in (3.3) are restricted by the condition $\mu_i' \geqq \mu_0$ and the minimum of (3.3) is attained by setting $\mu_i' = \mu_0 (i = 1, 2, \cdots, p_1)$. Since the result depends on the unknown integer p_1, we can obtain a lower bound by setting $p_1 = p$. Then using the symmetry of F we have

$$(3.4) \qquad P \geqq F^p(d).$$

The equation determining d is obtained by setting the right-hand member of (3.4) equal to P^* and is given by

$$(3.5) \qquad F(d) = (P^*)^{1/p}$$

It should be noted that (3.4) is independent of μ_0, σ and n_i. Hence with a table of the standard normal c.d.f. one can easily find the appropriate d which satisfies (3.5) and is to be used in rule (3.1) for any μ_0, any σ and any vector n_i.

The case when the normal populations have different but known variances and the standard is known is treated similarly The inequality defining the procedure for this problem, corresponding to (3.5), is

$$(3.6) \qquad \bar{x}_i \geqq \mu_0 - d\sigma_i/\sqrt{n_i}$$

and the equation determining d is exactly the same as (3.5)

Case B. *Common known variance* (μ_0 *unknown*) In this case n_0 independent observations are taken on the standard Π_0. Let $\bar{x}_0$ denote the mean of these n_0 observations and let σ^2 be the known common variance for all the $(p + 1)$ populations. Then the procedure is to select all those populations for which the relation

$$(3.7) \qquad \bar{x}_i \geqq \bar{x}_0 - d\sigma/\sqrt{n_i}$$

is satisfied. The equation determining d is obtained by the same argument as in Case A and, letting $f(x)$ denote the standard normal density, we obtain

$$(3.8) \qquad \int_{-\infty}^{\infty} \prod_{i=1}^{p} \left[F\left(u\sqrt{\frac{n_i}{n_0}} + d\right)\right] f(u)\, du = P^*.$$

For the special case $n_i = n (i = 0, 1, \cdots, p)$ this reduces to

$$(3.9) \qquad \int_{-\infty}^{\infty} F^p(u + d) f(u)\, du = P^*$$

Equation (3.9) is independent of σ. Hence a single two-way table of d-values for different values of P^* and p solves the problem for all values of σ when $n_i = n (i = 0, 1, \cdots, p)$. Tables of d-values satisfying (3.9) for several values of P^* are given in [2] for $p = 1$ (1) 10 and in [5] for $p = 1$ (1) 50. A short table, using only two decimals of the original four, is excerpted from [5] (see Table I). In the more general case when the populations have different but known variances the procedure is defined by

$$(3.10) \qquad \bar{x}_i \geqq \bar{x}_0 - d\sigma_i/\sqrt{n_i}$$

the best one is desired it can be chosen from the selected subset on the basis of economic or other considerations.

Although the title and discussion above use the phraseology "populations better than a standard" we shall actually be interested in selecting all populations as good as or better than the standard; for practical purposes the distinction is of minor importance since in most of the practical problems the parameters of interest can have any value in some interval and are very rarely equal.

To discuss confidence statements we consider first the problems below in which the better populations are the ones with the larger values of the main parameter of interest τ. After the experiment is performed, we can make with confidence P^* the joint statement that *for all populations which are eliminated the parameter value is less than that of the standard.* This joint confidence statement follows from the fact that in selecting a subset *containing all* populations as good as or better than a standard we are automatically eliminating a subset containing only populations worse than the standard. Hence this procedure can be used *to eliminate* those populations which are distinctly inferior to the standard.

For the case in which the better populations are defined to be the ones with the smaller values of τ, the statistical problem is identical and all the results and tables of this paper apply with the obvious modifications.

3. Location parameter—normal populations. We shall assume that populations $\Pi_1, \Pi_2, \cdots, \Pi_p$ with unknown means $\mu_1, \mu_2, \cdots, \mu_p$, respectively are given and that Π_0 is the standard or control, whose mean μ_0 may or may not be known. For clarity we shall discuss the various cases separately.

Case A. *Common known variance* (μ_0 *known*). From each of the p populations $\Pi_i (i = 1, 2, \cdots, p)$, n_i independent observations are taken. Let $\bar{x}_i$ denote the sample mean from Π_i and let σ^2 be the common known variance.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i (i = 1, 2, \cdots, p)$ for which

$$\bar{x}_i \geqq \mu_0 - d\sigma/\sqrt{n_i}\text{."} \tag{3.1}$$

To determine the value of d let p_1, p_2 denote the true number of populations with $\mu \geqq \mu_0$ and $\mu < \mu_0$, respectively, so that $p_1 + p_2 = p$. Then the probability P of retaining all the p_1 populations with $\mu \geqq \mu_0$ is given by

$$\begin{aligned} P &= \prod_{i=1}^{p_1} P\{\bar{x}_i' \geqq \mu_0 - d\sigma/\sqrt{n_i'}\} \\ &= \prod_{i=1}^{p_1} P\{\sqrt{n_i'}(\bar{x}_i' - \mu_i')/\sigma \geqq -d + \sqrt{n_i'}(\mu_0 - \mu_i')/\sigma\}, \end{aligned} \tag{3.2}$$

where primes refer to values associated with the p_1 populations for which $\mu \geqq \mu_0$. Hence

$$P = \prod_{i=1}^{p_1} \{1 - F(-d + \sqrt{n_i'}(\mu_0 - \mu_i')/\sigma)\} \tag{3.3}$$

where $F(x)$ refers to the standard normal cumulative distribution function. The μ_i' in (3.3) are restricted by the condition $\mu_i' \geqq \mu_0$ and the minimum of (3.3) is attained by setting $\mu_i' = \mu_0 (i = 1, 2, \cdots, p_1)$. Since the result depends on the unknown integer p_1, we can obtain a lower bound by setting $p_1 = p$. Then using the symmetry of F we have

$$P \geqq F^p(d). \tag{3.4}$$

The equation determining d is obtained by setting the right-hand member of (3.4) equal to P^* and is given by

$$F(d) = (P^*)^{1/p}. \tag{3.5}$$

It should be noted that (3.4) is independent of μ_0, τ and n_i. Hence with a table of the standard normal c.d.f. one can easily find the appropriate d which satisfies (3.5) and is to be used in rule (3.1) for any μ_0, any σ and any vector n_i.

The case when the normal populations have different but known variances and the standard is known is treated similarly The inequality defining the procedure for this problem, corresponding to (3.5), is

$$\bar{x}_i \geqq \mu_0 - d\sigma_i/\sqrt{n_i} \tag{3.6}$$

and the equation determining d is exactly the same as (3.5).

Case B. *Common known variance* (μ_0 *unknown*). In this case n_0 independent observations are taken on the standard Π_0 Let $\bar{x}_0$ denote the mean of these n_0 observations and let σ^2 be the known common variance for all the $(p + 1)$ populations. Then the procedure is to select all those populations for which the relation

$$\bar{x}_i \geqq \bar{x}_0 - d\sigma/\sqrt{n_i} \tag{3.7}$$

is satisfied. The equation determining d is obtained by the same argument as in Case A and, letting $f(x)$ denote the standard normal density, we obtain

$$\int_{-\infty}^{\infty} \prod_{i=1}^{p} \left[F\left(u \sqrt{\frac{n_i}{n_0}} + d \right) \right] f(u)\, du = P^* \tag{3.8}$$

For the special case $n_i = n (i = 0, 1, \cdots, p)$ this reduces to

$$\int_{-\infty}^{\infty} F^p(u + d) f(u)\, du = P^* \tag{3.9}$$

Equation (3.9) is independent of σ. Hence a single two-way table of d-values for different values of P^* and p solves the problem for all values of σ when $n_i = n (i = 0, 1, \cdots, p)$. Tables of d-values satisfying (3.9) for several values of P^* are given in [2] for $p = 1$ (1) 10 and in [5] for $p = 1$ (1) 50. A short table, using only two decimals of the original four, is excerpted from [5] (see Table I). In the more general case when the populations have different but known variances the procedure is defined by

$$\bar{x}_i \geqq \bar{x}_0 - d\sigma_i/\sqrt{n_i} \tag{3.10}$$

TABLE I[a]

Table of d-values satisfying (3.9) and used in the procedure defined by (3.7)

p	P^*			
	.75	.90	.95	.99
1	0.95	1.81	2.33	3.29
2	1.43	2.23	2.71	3.62
3	1.68	2.45	2.92	3.80
4	1.85	2.60	3.06	3.92
5	1.97	2.71	3.16	4.01
6	2.06	2.80	3.24	4.09
7	2.14	2.87	3.31	4.15
8	2.21	2.93	3.37	4.20
9	2.26	2.98	3.42	4.25
10	2.31	3.03	3.46	4.29
15	2.50	3.20	3.63	4.44
20	2.62	3.32	3.74	4.54
30	2.79	3.48	3.89	4.68
40	2.90	3.58	4.00	4.78
50	2.99	3.67	4.08	4.85

[a] For a more complete table see [5].

and the equation determining d is

$$(3.11) \qquad \int_{-\infty}^{\infty} \prod_{i=1}^{p} \left[F\left(u \frac{\sigma_0}{\sigma_i} \sqrt{\frac{n_i}{n_0}} + d \right) \right] f(u)\, du = P^*.$$

this reduces to (3.9) in the case when $\sigma_i/\sqrt{n_i} =$ constant $(i = 0, 1, \cdots, p)$.

Case C. *Common unknown variance* (μ_0 *known*). As in Case A, n_i observations are taken only on the p populations $\Pi_i(i = 1, 2, \cdots, p)$. Let s_ν^2 denote the pooled estimate of σ^2 based on $\nu = \sum_{i=1}^{p} (n_i - 1)$ degrees of freedom ($n_i > 1$ for at least one i). Then the procedure is to select those and only those populations Π_i for which

$$(3.12) \qquad \bar{x}_i \geqq \mu_0 - ds_\nu/\sqrt{n_i}\,.$$

The equation determining d is

$$(3.13) \qquad \int_0^{\infty} F^p(yd)q_\nu(y)\, dy = P^*,$$

where $q_\nu(y)$ is the density of $y = s_\nu/\sigma = \chi_\nu/\sqrt{\nu}$. This result holds for any μ_0 and depends on n_i only through the value of ν.

Case D. *Common unknown variance* (μ_0 *unknown*). In this case n_i observations are taken on all the populations $\Pi_i(i = 0, 1, \cdots, p)$ and the pooled estimate s_ν^2 of σ^2 is based on $\nu = \sum_{i=0}^{p} (n_i - 1)$ d.f. ($n_i > 1$ for at least one i).

The inequality defining the procedure is

$$(3.14) \qquad \bar{x}_i \geqq \bar{x}_0 - ds_\nu/\sqrt{n_i}\,.$$

The equation determining d is

$$\int_0^\infty \int_{-\infty}^\infty \left[\prod_{i=1}^{p} F\left(u\sqrt{\frac{n_i}{n_0}} + yd\right)\right] f(u)q_\nu(y)\, du\, dy = P^*. \tag{3.15}$$

For $n_i = n(i = 0, 1, \cdots, p)$ this reduces to

$$\int_0^\infty \int_{-\infty}^\infty F^p(u + yd)f(u)q_\nu(y)\, du\, dy = P^*. \tag{3.16}$$

Methods for evaluating this double integral and tables of d-values for selected values of P^*, p and ν are given in [6] and values of $d/\sqrt{2}$ for other values of p and ν are given in [3].

4. Scale parameter—gamma or chi-square populations. *In this section it will* be more natural to define the population Π_i as better than Π_0 if the scale parameter $\theta_i < \theta_0$.

Case A. θ_0 *known.* We assume that the population $\Pi_i (i = 1, 2, \cdots, p)$ has the density

$$\frac{1}{\Gamma\left(\frac{\alpha_i}{2}\right)} \frac{1}{\theta_i^{\alpha_i/2}} x^{\frac{\alpha_i}{2}-1} e^{-x/\theta_i}. \tag{4.1}$$

If $x_{ij}(j = 1, 2, \cdots, n_i)$ are the n_i observations on Π_i, then $t_i = \sum_{j=1}^{n_i} x_{ij}$ has the density (4.1) with α_i replaced by $\nu_i = n_i\alpha_i$ and the procedure is as follows.

Procedure: "Retain in the selected subset only those populations

$$\Pi_i(i = 1, 2, \cdots, p)$$

for which

$$\frac{t_i}{\nu_i} \leqq (1 + d)\theta_0." \tag{4.2}$$

Let q_1 and q_2 denote the number of populations with $\theta \leqq \theta_0$ and $\theta > \theta_0$, respectively, so that $q_1 + q_2 = p$. The probability P of a correct decision is given by

$$P = \prod_{i=1}^{q_1} P\left\{\frac{t_i'}{\theta_i'} \leqq (1 + d)\frac{\theta_0\nu_i'}{\theta_i'}\right\}, \tag{4.3}$$

where primes refer to the q_1 populations with $\theta \leqq \theta_0$. Hence,

$$P = \prod_{i=1}^{q_1} G_{\nu_i'}\left[(1 + d)\frac{\theta_0\nu_i'}{\theta_i'}\right], \tag{4.4}$$

where $G_{\nu_i}(x)$ is the c.d.f. of the gamma density in (4.1) with α_i replaced by ν_i and $\theta_i = 1$. A lower bound to this probability is obtained by setting $\theta_i = \theta_0$

and $q_1 = p$ so that the equation determining d can be written in the form

$$(4.5)\qquad \prod_{i=1}^{p} \left\{ \frac{1}{\Gamma\left(\frac{\nu_i}{2}\right)} \int_0^{\nu_i(1+d)} u^{\frac{\nu_i}{2}-1} e^{-u}\, du \right\} = P^*.$$

For $\nu_i = \nu (i = 1, 2, \cdots, p)$ this is easily solved with the help of a table of the c.d.f. of $\gamma_\nu = \frac{1}{2}\chi_\nu^2$ with ν degrees of freedom.

Application to normal populations. If $\theta_i = 2\sigma_i^2 (i = 0, 1, \cdots, p)$ are the scale parameters for the $(p + 1)$ normal populations and $x_{ij} (j = 1, 2, \cdots, n_i)$ are the n_i observations on the population Π_i with the mean μ_i (known), then we retain the population Π_i in the selected subset if

$$(4.6)\qquad s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \leqq 2(1 + d)\sigma_0^2.$$

The equation determining the d in (4.6) is the same as (4.5) with ν_i replaced by n_i.

If the means μ_i are unknown and $n_i > 1 (i = 1, 2, \cdots, p)$, then in (4.6) we use the sample mean $\bar{x}_i$ in place of μ_i and $n_i - 1$ in place of n_i. The equation determining d is again (4.5) with $\nu_i = n_i - 1$.

Transformation: If we apply the transformation [1]

$$(4.7)\qquad y_i = \ln\left(\frac{t_i}{\nu_i}\right) \qquad (i = 1, 2, \cdots, p),$$

then the procedure (4.2) of this section can be put in the form

$$(4.8)\qquad y_i \leqq \ln\left(\frac{\theta_0}{2}\right) + d_1,$$

where

$$(4.9)\qquad d_1 = \ln\,[2(1 + d)].$$

Then using the normal approximation and the same argument as before, the approximate equation determining d_1 is

$$(4.10)\qquad \prod_{i=1}^{p} \left\{ F\left(d_1 \sqrt{\frac{\nu_i}{2}}\right) \right\} = P^*.$$

For $\nu_i = \nu (i = 1, 2, \cdots, p)$ this gives an equation similar to (3.5). For the application to normal populations the equation corresponding to (4.8) is

$$(4.11)\qquad \ln s_i^2 \leqq \ln \sigma_0^2 + d_1,$$

where d_1 is determined by (4.10) with $\nu_i = n_i$ or $n_i - 1$ according as the means μ_i are or are not known.

Case B. *θ_0 unknown.* The assumptions are the same as in Case A except that n_0 observations, viz., $x_{01}, x_{02}, \cdots, x_{0n_0}$ are taken on Π_0. The inequality de-

fining the procedure and corresponding to (4.2) is

$$\frac{t_i}{\nu_i} \leqq (1 + d) \frac{t_0}{\nu_0}, \tag{4.12}$$

where $t_0 = \sum_{j=1}^{n_0} x_{0j}$ and $\nu_0 = n_0 \alpha_0$. The equation determining d is obtained as before and is given by

$$\int_0^\infty \left[\prod_{i=1}^p \int_0^{\nu_i t(1+d)/\nu_0} \frac{u^{\frac{\nu_i}{2}-1} e^{-\frac{u}{2}}}{\Gamma(n_i/2)} du \right] \frac{t^{\frac{\nu_0}{2}-1} e^{-\frac{t}{2}}}{\Gamma(\nu_0/2)} dt. \tag{4.13}$$

Application to normal populations. For the case where the means are known, the rule takes the form

$$\frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \leqq \frac{(1 + d)}{n_0} \sum_{j=1}^{n_0} (x_{0j} - \mu_0)^2, \tag{4.14}$$

where d is given by (4.13) with $\nu_i = n_i$. *If* μ_i*'s are not known and*

$$n_i > 1 (i = 0, 1, \cdots, p),$$

then the rule is the same as (4 14) with μ_i and n_i replaced by $\bar{x}_i$ and $n_i - 1$, *respectively. The equation determining* d *is again (4.13) with* $\nu_i = n_i - 1$.

Transformation: Using the transformation (4.7), we put the inequality defining the rule as

$$y_i \leqq y_0 + d_2 \tag{4.15}$$

The approximate equation determining d_2 is

$$\int_{-\infty}^\infty \left[\prod_{i=1}^p F\left(u \sqrt{\frac{n_i}{n_0}} + d_2 \sqrt{\frac{n_i}{2}} \right) \right] f(u)\, du = P^*, \tag{4.16}$$

which for $n_i = n (i = 0, 1, \cdots, p)$ is of the same form as (3.9).

5. Binomial populations.

Case A. *Known standard.* It is assumed that $p + 1$ binomial populations Π_i with parameters $\theta_i (i = 0, 1, \cdots, p)$ are given where θ_0 is the known value of the probability of a unit being defective in the standard population Π_0. Again n_i independent observations are taken from each population

$$\Pi_i (i = 1, 2, \cdots, p)$$

Since θ_i is the probability of a unit being defective, we define Π_i to be better than Π_0 when $\theta_i < \theta_0$. Let x_i denote the number of defectives observed in the sample of n_i observations from $\Pi_i (i = 1, 2, \cdots, p)$.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i (i = 1, 2, \cdots, p)$ for which

$$\frac{1}{n_i} x_i \leqq \theta_0 + d \sqrt{\frac{\theta_0(1 - \theta_0)}{n_i}}." \tag{5.1}$$

Let q_1, q_2, be defined as in Sec. 4; let $[m_i(d)]$ denote the largest integer in

$$m_i(d) = n_i\theta_0 + d\sqrt{n_i\theta_0(1-\theta_0)} \qquad (i = 1, 2, \cdots, p). \tag{5.2}$$

The probability P of retaining all the q_1 populations with $\theta \leqq \theta_0$ is given by

$$P = \prod_{i=1}^{q_1}\left[\sum_{j=0}^{[m_i(d)]} C_j^{n_i}\theta_i^j(1-\theta_i)^{n_i-j}\right]. \tag{5.3}$$

A lower bound is obtained by setting $\theta_i = \theta_0(i = 1, 2, \cdots, p)$ and $q_1 = p$. The fact that $\theta_i = \theta_0$ gives a lower bound can be shown by writing the binomial sum as an incomplete Beta function. Hence the inequality determining d becomes

$$\prod_{i=1}^{p}\left[\sum_{j=0}^{[m_i(d)]} C_j^{n_i}\theta_0^j(1-\theta_0)^{n_i-j}\right] \geqq P^*, \tag{5.4}$$

and the solution is the smallest value of d satisfying (5.4). If $n_i = n$ then

$$[m_i(d)] = [m(d)]$$

and (5.4) reduces to

$$\sum_{j=0}^{[m(d)]} C_j^n\theta_0^j(1-\theta_0)^{n-j} \geqq (P^*)^{1/p}. \tag{5.5}$$

This is easily solved by consulting a table of cumulative binomial probabilities.

For large values of n, (large enough for the normal approximation to give good results) the inequality determining d can be approximated by the simple equation

$$F(d) = (P^*)^{1/p}, \tag{5.6}$$

where F is the standard normal c.d.f. This equation is independent of n_i and is much easier to solve than (5.4).

Case B. *Unknown standard.* The assumptions are the same as in Case A except that n_0 observations are taken on the standard population Π_0. Let x_0 be the number of defectives among n_0.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i(i = 1, 2, \cdots, p)$ for which

$$\frac{1}{n_i}x_i \leqq \frac{1}{n_0}x_0 + \frac{d}{2}\sqrt{\frac{1}{n_i}+\frac{1}{n_0}}." \tag{5.7}$$

The probability P of retaining all the q_1 populations with $\theta \leqq \theta_0$ attains a minimum when $\theta_i = \theta(i = 0, 1, \cdots, p)$ and $q_1 = p$ and is given by

$$P(\theta, d) = \sum_{y=0}^{n_0}\prod_{i=1}^{p}\left[\sum_{j=0}^{[m_i(y,d)]} C_j^{n_i}\theta^j(1-\theta)^{n_i-j}\right] C_y^{n_0}\theta^y(1-\theta)^{n_0-y}, \tag{5.8}$$

where $[m_i(y, d)]$ is the largest integer contained in

$$m_i(y, d) = \frac{n_i}{n_0} y + \frac{dn_i}{2} \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}. \tag{5.9}$$

Then the desired value of d for (5.7) is the smallest number for which

$$\min_{0 \leq \theta \leq 1} P(\theta, d) \geq P^*. \tag{5.10}$$

Since, except for very small n_i or very large p, the minimum occurs near $\theta = \frac{1}{2}$, we can obtain an approximate solution for d by finding the smallest number for which

$$P(\tfrac{1}{2}, d) \geq P^*. \tag{5.11}$$

A simpler approximate solution, which gives good results when the n_i are not too small and p is not too large, is the normal approximation obtained under the assumption that $\theta_i = \frac{1}{2} (i = 0, 1, \cdots, p)$. Then from (5.7) we obtain for the approximate equation determining d

$$\int_{-\infty}^{\infty} \left[\prod_{i=1}^{p} F\left(u \sqrt{\frac{n_i}{n_0}} + d \sqrt{1 + \frac{n_i}{n_0}} \right) \right] f(u) \, du = P^*. \tag{5.12}$$

For $n_i = n (i = 0, 1, \cdots, p)$ the rule (5.7) can be written as

$$x_i \leq x_0 + d', \tag{5.13}$$

where $d' = d\sqrt{n/2}$. In carrying out the rule we can assume that d' is an *integer*. The desired value of d' is the smallest integer for which

$$\min_{0 \leq \theta \leq 1} \left\{ \sum_{y=0}^{n} \left[\sum_{j=0}^{y+d'} C_j^n \theta^j (1 - \theta)^{n-j} \right]^p C_y^n \theta^y (1 - \theta)^{n-y} \right\} \geq P^*. \tag{5.14}$$

Then (5.12) can be written in the form

$$\int_{-\infty}^{\infty} F^p(u + \hat{d}) f(u) \, du = P^* \tag{5.15}$$

and the relation between d' and $\hat{d}$, using a continuity correction, is

$$d' = d \sqrt{\frac{n}{2}} = \left\{ \frac{\hat{d}\sqrt{n} - 1}{2} \right\} \tag{5.16}$$

where $\{x\}$ is the smallest integer greater than or equal to x.

Transformation: It may be desirable to solve the binomial problem by using an arc sine transformation and converting it into one involving the location parameter of the normal distributions. For example, for the Case B above with $n_i = n (i = 0, 1, \cdots, p)$ if we use the arc sine transformation as given in [4],

the inequality defining the procedure is

$$\text{(5.17)} \quad \begin{aligned} \arcsin \sqrt{\frac{x_i}{n+1}} + \arcsin \sqrt{\frac{x_i+1}{n+1}} \leqq \arcsin \sqrt{\frac{x_0}{n+1}} \\ + \arcsin \sqrt{\frac{x_0+1}{n+1}} + \frac{d\sqrt{2}}{\sqrt{2n+1}}, \end{aligned}$$

where the approximate equation determining d is the same as (3.9) so that Table I is applicable here also.

REFERENCES

[1] M. S. BARTLETT AND D. G. KENDALL, "The statistical analysis of variance-heterogeneity and the logarithmic transformation," *J. Roy. Stat. Soc. Suppl.*, Vol. 8 (1946), pp. 128–138.

[2] R. E. BECHHOFER, "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 16–29.

[3] C. W. DUNNETT, "A multiple comparison procedure for comparing several treatments with a control," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 1096–1121.

[4] M. F. FREEMAN AND J. W. TUKEY, "Transformations related to the angular and the square root," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 607–611.

[5] S. S. GUPTA, "On a decision rule for a problem in ranking means," Mimeograph Series No. 150, May, 1956, Institute of Statistics, University of North Carolina, Chapel Hill, N. C.

[6] S. S. GUPTA AND M. SOBEL, "On a statistic which arises in selection and ranking problems," submitted for publication in *Ann. Math. Stat.*

[7] S. S. GUPTA AND M. SOBEL, "A procedure for selecting a subset containing the best of k binomial populations," to be published.

[8] E. PAULSON, "On the comparison of several experimental categories with a control," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 239–246.

SOME PROBLEMS OF SIMULTANEOUS MINIMAX ESTIMATION

By Stanisław Trybuła

Institute of Mathematics, Polish Academy of Sciences, Wrocław

1. Summary. In this paper, we give minimax estimates of the parameters of the multivariate hypergeometric distribution and of the multinomial distribution, and of some parameters of an unspecified distribution with known range. We use as loss a weighted linear combination of squared differences between the true and the estimated values of the parameters. Some properties of the minimax estimates obtained are discussed

2. Introduction. For our purpose, it is sufficient to define the estimation problem in a fixed sample size experiment as follows ([3], [4]). The random variable X is distributed in the space $\mathfrak{X}$ according to the distribution F belonging to the family $\mathfrak{F}$. We want to estimate $\omega(F)$ where ω is a function, the values of which belong to some space Ω, defined on $\mathfrak{F}$. (In the following we assume that X and $\omega(F)$ are vector valued.) An estimate is a statistic $f(X)$ having values in Ω. The nonnegative function $L[\omega(F), f(x)]$ is the loss resulting if, when F obtains, the estimate $f(x)$ is made. Define the risk by

$$R(f, F) = E\{L[\omega(F), f(X)]|F\} \tag{1}$$

and call $v(f) = \sup_{F \in \mathfrak{F}} R(f, F)$ the guaranteed value for the estimate f. We seek the minimax estimate f^0, that is, the estimate whose guaranteed value is minimal. Obviously, such an estimate does not always exist. It is our aim to derive minimax estimates in some specific problems.

3. Problem 1. In practice, we often meet the following situation. A lot consisting of N units of a product has been produced. The units are classified into l categories, the ith category containing U_i units ($i = 1, \cdots, l$). A sample of size n is taken from the lot in which $k_1, \cdots, k_l$ units of categories $1, \cdots, l$ are observed. The problem is to estimate $U_1, \cdots, U_l$.

This leads to the estimation of the parameter $U = (U_1, \cdots, U_l)$ of a multivariate hypergeometric distribution. Thus, let

$$P(X_1 = k_1, \cdots, X_l = k_l) = \frac{\binom{U_1}{k_1} \cdots \binom{U_l}{k_l}}{\binom{N}{n}}. \tag{2}$$

It is known that,

$$m_i = E(X_i \mid U) = n\frac{U_i}{N}, \tag{3}$$

Received April 12, 1957; revised July 11, 1957.

$$\sigma_i^2 = E\{[X_i - E(X_i \mid U)]^2 \mid U\} = \frac{n(N-n)}{N^2(N-1)} U_i(N - U_i). \tag{4}$$

Suppose that the loss is

$$L(U, f) = \sum_{i=1}^{l} c_i[f_i(X) - U_i]^2 \qquad (c_i \geqq 0), \tag{5}$$

where $f = (f_1, \cdots, f_l)$ is the estimate of U and $X = (X_1, \cdots, X_n)$ is the sample. The risk is then

$$R(f, U) = E[L(U, f) \mid U] = E\left\{\sum_{i=1}^{l} c_i[f_i(X) - U_i]^2 \mid U\right\}. \tag{6}$$

If we study estimates of the form

$$f_i(X) = aX_i + b_i \qquad (i = 1, \cdots, l),$$

then

$$\begin{aligned} R(f, U) &= \sum_{i=1}^{l} c_i E\{[aX_i + b_i - U_i]^2 \mid U\} \\ &= \sum_{i=1}^{l} c_i[(am_i + b_i - U_i)^2 + a^2\sigma_i^2] \qquad (7) \\ &= \sum_{i=1}^{l} c_i \left[\left(a\frac{nU_i}{N} + b_i - U_i\right)^2 + a^2 \frac{n(N-n)}{N^2(N-1)} U_i(N - U_i)\right]. \end{aligned}$$

Let the constant a assume a value such that the terms quadratic in U vanish. For this, it suffices to put

$$a = \frac{N}{n + \sqrt{n\frac{N-n}{N-1}}}.$$

If, moreover, we put

$$b_i = \frac{s_i N \sqrt{n\frac{N-n}{n-1}}}{n + \sqrt{n\frac{N-n}{N-1}}},$$

then (7) may be written

$$R(f, U) = \frac{nN\frac{N-n}{N-1}}{\left(n + \sqrt{n\frac{N-n}{N-1}}\right)^2} \sum_{i=1}^{l} c_i[Ns_i^2 + (1 - 2s_i)U_i]. \tag{8}$$

Without loss of generality, we may assume $c_1 \geqq c_2 \geqq \cdots \geqq c_l \geqq 0$. For the present, assume also that $c_2 \neq 0$. Let l_0 be the greatest index i for which $c_i \neq 0$

and let

(9) $$L = \max_s \left[s \leqq l_0, \sum_{i=1}^{s} 1/c_i > \frac{s-2}{c_s} \right].$$

The above assumptions being satisfied, we prove the following lemma:

If $L \leqq l$ then

(10) $$\delta = \frac{L-2}{\sum_{j=1}^{L} 1/c_j} \geqq c_i \text{ for } i = L+1, L+2, \cdots, l.$$

PROOF. First, observe that a proof of the inequality is necessary only for $i = L + 1$. If $c_{L+1} = 0$, then the lemma obviously holds. If $c_{L+1} \neq 0$, it follows from the definition of L that

$$L - 1 \geqq c_{L+1} \sum_{j=1}^{L+1} \frac{1}{c_j} = 1 + c_{L+1} \sum_{j=1}^{L} 1/c_j .$$

The lemma is a direct consequence of this inequality.

Now put

(11) $$s_i = \begin{cases} \frac{1}{2}\left(1 - \frac{\delta}{c_i}\right), & \text{when } i \leqq L, \\ 0, & \text{when } i > L. \end{cases}$$

Observe that $i \leqq L$, $0 < s_i \leqq \frac{1}{2}$. We shall show that the estimate

$$f^0 = (f_1^0, f_2^0, \cdots, f_l^0),$$

where

(12) $$f_i^0(X) = N \frac{X_i + s_i \sqrt{n \frac{N-n}{N-1}}}{n + \sqrt{n \frac{N-n}{N-1}}},$$

is the minimax estimate sought.

From (8) and (11) we have

(13) $$R(f^0, U) = \frac{Nn \frac{N-n}{N-1}}{\left(n + \sqrt{n \frac{N-n}{N-1}}\right)^2} \left\{ \sum_{i=1}^{L} \left[c_i \frac{N}{4} \left(1 - \frac{\delta}{c_i}\right)^2 + \delta U_i \right] + \sum_{i=L+1}^{l} c_i U_i \right\}.$$

Observe that for

(14) $$U_{L+1} = U_{L+2} = \cdots = U_l = 0,$$

$R(f^0, U) = c$, where c is a constant. By the lemma, $R(f^0, U) \leqq c$. Thus, by theorem 2.1 of [4], it is sufficient to prove that a distribution of the random variable U exists which satisfies (14) and for which f^0 is the Bayes estimate.

We seek for such a distribution among those of the form

$$(15) \qquad P(U_{L+1} = \cdots = U_l = 0) = 1$$

$$(16) \qquad P(U_1 = u_1, \cdots, U_L = u_L) = C \frac{\Gamma(a_1 + u_1) \cdots \Gamma(a_L + u_L)}{u_1! \cdots u_L!}.$$

Let

$$(17) \qquad r(f, P) = E[R(f, U)] = \sum_{i=1}^{l} c_i E\{E[f_i(X) - U_i]^2 \mid U\}.$$

It follows from (15) that the expected risk does not depend on f_i if $k_j \neq 0$ for at least one $j > L$. Thus, any estimate which minimizes (17) throughout the region $k_{L+1} = k_{L+2} = \cdots = k_l = 0$ is a Bayes estimate. Now, if

$$k_{L+1} = k_{L+2} = \cdots = k_l = 0$$

then, as is well-known, the expression (17) attains its minimum value for

$$(18) \qquad \begin{aligned} &f_i(k_1, \cdots, k_L, 0, \cdots, 0) \\ &= E(U_i \mid X_1 = k_1, \cdots, X_L = k_L; X_{L+1} = \cdots = X_l = 0) \\ &= \begin{cases} 0 & \text{for } i > L; \\ \dfrac{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1 \geq k_1, \cdots, u_L \geq k_L}} u_i \prod_{j=1}^{L} \binom{u_j}{k_j} \frac{\Gamma(a_j + u_j)}{u_j!}}{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1 \geq k_1, \cdots, u_L \geq k_L}} \prod_{j=1}^{L} \binom{u_j}{k_j} \frac{\Gamma(a_j + u_j)}{u_j!}} & \text{otherwise.} \end{cases} \end{aligned}$$

The second part of Eq. (18) reduces to

$$\frac{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1 \geq k_1, \cdots, u_L \geq k_L}} u_i \prod_{j=1}^{L} \frac{\Gamma(a_j + u_j)}{(u_j - k_j)!}}{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1 \geq k_1, \cdots, u_L \geq k_L}} \prod_{j=1}^{L} \frac{\Gamma(a_j + u_j)}{(u_j - k_j)!}}$$

$$= \frac{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1 \geq 0, \cdots, v_L \geq 0}} [(a_i + k_i + v_i) - a_i] \prod_{j=1}^{L} \frac{\Gamma(a_j + k_j + v_j)}{v_j!}}{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1 \geq 0, \cdots, v_L \geq 0}} \prod_{j=1}^{L} \frac{\Gamma(a_j + k_j + v_j)}{v_j!}}$$

$$= \frac{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1 \geq 0, \cdots, v_L \geq 0}} \Gamma(a_i + k_i + v_i + 1) \prod_{\substack{j=1 \\ j \neq i}}^{L} \frac{\Gamma(a_j + k_j + v_j)}{v_j!}}{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1 \geq 0, \cdots, v_L \geq 0}} \prod_{j=1}^{L} \frac{\Gamma(a_i + k_j + v_j)}{v_j!}} - a_i$$

$$= \frac{L_i}{M_i} - a_i.$$

Observe that

$$\sum_{\substack{v_1+\cdots+v_L=N-n\\ v_1\geq 0,\cdots,v_L\geq 0}} \frac{(N-n)!}{v_1!\cdots v_L!} \frac{\Gamma(b_1+v_1)\cdots\Gamma(b_L+v_L)}{\Gamma\left(N-n+\sum_{j=1}^{L} b_j\right)}$$

(19)
$$= \int\cdots\int_{\substack{p_1+\cdots+p_L=1\\ p_1\geq 0,\cdots,p_L\geq 0}} p_1^{b_1-1}\cdots p_L^{b_L-1} \sum_{\substack{v_1+\cdots+v_L=N-n\\ v_1\geq 0,\cdots,v_L\geq 0}} \frac{(N-n)!}{v_1!\cdots v_L!} \cdot p_1^{v_1}\cdots p_L^{v_L}\,dp_1\cdots dp_L$$

$$= \int\cdots\int_{\substack{p_1+\cdots+p_L=1\\ p_1\geq 0,\cdots,p_L\geq 0}} p_1^{b_1-1}\cdots p_L^{b_L-1}\,dp_1\cdots dp_L = \frac{\Gamma(b_1)\cdots\Gamma(b_L)}{\Gamma\left(\sum_{i=1}^{L} b_i\right)}.$$

Applying (19) to L_i with $b_j = a_j + k_j$ for $j \neq i$ and

$$b_i = a_i + k_i + 1 (i = 1, \cdots, L)$$

and to M_i with $b_j = a_j + k_j$ we obtain

(20)
$$\frac{L_i}{M_i} - a_i = \frac{(a_i + k_i)\left(N + \sum_{j=1}^{L} a_j\right)}{n + \sum_{j=1}^{L} a_j} - a_i = \frac{\left(N + \sum_{j=1}^{L} a_j\right) k_i + (N-n)a_i}{n + \sum_{j=1}^{L} a_j} = f_i^0(k_1, \cdots, k_L, 0, \cdots, 0)$$

for

$$a_i = s_i \frac{N\sqrt{n\dfrac{N-n}{N-1}}}{N - n - \sqrt{n\dfrac{N-n}{N-1}}}.$$

Thus f^0 is minimax whenever $a_i > 0$; that is, when $N > n + 1$. For $N = n$ this result is immediate. For $N = n + 1$ it is a consequence of the fact that f^0 is Bayes for the a priori distribution of U defined by

$$P(U_{L+1} = \cdots = U_l = 0) = 1,$$

$$P(U_1 = u_1, \cdots, U_L = u_L) = \frac{N!}{u_1!\cdots u_L!} s_1^{u_1}\cdots s_L^{u_L}.$$

Up to this point, we have assumed $c_2 \neq 0$. Consider now the remaining cases. If all $c_i = 0$ then, obviously, every estimate is minimax. If alone $c_1 \neq 0$ then the problem may be considered as that of finding a minimax estimate of th

parameter U in a one-dimensional hypergeometric distribution for the loss $L = (f - U)^2$. In this case, the formula for the minimax estimate is (see [4])

$$(21) \qquad f(X) = N \frac{X + \frac{1}{2}\sqrt{n \frac{N-n}{N-1}}}{n + \sqrt{n \frac{N-n}{N-1}}}.$$

It is easy to verify that the estimate (12) satisfies the condition

$$\sum_{i=1}^{L} f_i^0 = N.$$

Observe that we are actually dealing with only $l - 1$ independent parameters since one parameter, say U_l, may be computed from

$$(22) \qquad U_1 + \cdots + U_l = N.$$

If we consider the problem of finding a minimax estimate for $U_1, \cdots, U_{l-1}$ under the loss

$$(23) \qquad \bar{L}(U, f) = \sum_{i=1}^{l-1} \bar{c}_i(f_i - U_i),$$

the same estimate as above for $U_1, \cdots, U_{l-1}$ results as is seen by identifying c_i in the above with $\bar{c}_i (i = 1, \cdots, l - 1)$ and putting $c_l = 0$.

In solving our problem, we have restricted ourselves to the case $c_i \geqq 0$. If, however, some $c_j < 0$ then for $f_j \to \pm\infty$ the loss tends to $-\infty$ and, consequently, the problem becomes trivial.

In the special case $c_1 = c_2 = \cdots = c_l > 0$, formula (12) takes the form

$$(24) \qquad f_i^0(x) = N \frac{X_i + \frac{1}{l}\sqrt{n \frac{N-n}{N-1}}}{n + \sqrt{n \frac{N-n}{N-1}}}.$$

4. Corollaries for the multinomial case. For $N \to \infty$, the distribution of X converges to the multinomial distribution defined by

$$P(X_1 = k_1, \cdots, X_l = k_l) = \frac{n!}{k_1! \cdots k_l!} p_1^{k_1} \cdots p_l^{k_l},$$

$$0 \leqq k_i, 0 \leqq p_i \leqq 1, i = 1, \cdots, l'$$

$$\sum_1^l k_i = n;$$

and

$$\lim_{N\to\infty} \frac{f_i^0(X)}{N} = \frac{X_i + s_i\sqrt{n}}{n + \sqrt{n}} = g_i^0(x).$$

We shall prove[1] that $g^0 = (g_1^0, \cdots, g_l^0)$ is really a minimax estimate of the parameter $p = (p_1, \cdots, p_l)$ for the loss

$$L(g, p) = \sum_{i=1}^{l} c_i(g_i - p_i)^2, \qquad c_i \geq 0. \tag{25}$$

When L, δ and s_i are defined by (9), (10), and (11), respectively, the loss is

$$R(g^0, p) = E\{L(g^0, p) \mid p\} = \frac{1}{(\sqrt{n} + 1)^2}\left[\sum_{i=1}^{L} (c_i s_i^2 + \delta p_i) + \sum_{i=L+1}^{l} c_i p_i\right], \tag{26}$$

which for $p_{L+1} = \cdots = p_l = 0$ is constant and, by the lemma of Sec. 2, maximum in p.

As is easy to verify, g^0 is Bayes for the a priori distribution $G(p)$ defined by

$$dG(p) = \begin{cases} Cp_1^{\sqrt{n}s_1 - 1} \cdots p_L^{\sqrt{n}s_L - 1}, & \text{when } p_{L+1} = \cdots = p_l = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

By theorem 2.1 of [4] it follows that g^0 is minimax.

For $c_1 = \cdots = c_l > 0$, $s_i = 1/l$ and the minimax estimate g^0 takes the form

$$g_i^0(X) = \frac{X_i + \frac{1}{l}\sqrt{n}}{n + \sqrt{n}}.$$

This case was previously solved by H. Steinhaus in [5].

5. Problem 2. We shall prove the following theorem:

THEOREM. *Let X be a random variable distributed according to the unknown distribution F on the measurable space A. Let $g_1, \cdots, g_m$ be such bounded measurable functions on A that there exist two points x' x'' ε A such that each of these functions attains its minimum in x' and its maximum in x''. Let $X_1, \cdots, X_n$ be a random sample from F, and let $\lambda_i = E[g_i(X)]$. If the loss is given by*

$$L(f, \lambda) = \sum_{i=1}^{m} c_i(f_i - \lambda_i)^2, \tag{29}$$

where $f = (f_1, \cdots, f_m)$ is an estimate of $\lambda = (\lambda_1, \cdots, \lambda_m)$, then the minimax estimate of λ is given by

$$f_i^0(X_1, \cdots, X_n) = \frac{\sum_{j=1}^{n} g_i(X_j)}{n + \sqrt{n}} + \frac{s_i}{\sqrt{n} + 1} \tag{30}$$

(s_i is the arithmetic mean of the maximum and minimum values of $g_i(x)$).

PROOF. If $f_i(X_1, \cdots, X_n) = a \sum_{j=1}^{n} g_i(X_j) + b_i$, then the risk may be

[1] While this paper was being written, Joseph Dubay communicated to me a result simil to this but not in its full generality.

written

$$(31)\quad \begin{aligned} R(\bar{f}, F) &= E\left[\sum_{i=1}^{m} c_i(f_i - \lambda_i)^2 \mid F\right] = \sum_{i=1}^{m} c_i E\left\{\left[a \sum_{j=1}^{n} g_i(X_j) + b_i - \lambda_i\right]^2 \middle| F\right\} \\ &= \sum_{i=1}^{m} c_i\{[(1 - an)\lambda_i - b_i]^2 + na^2 E\{[g_i(X) - \lambda_i]^2 \mid F\}\}. \end{aligned}$$

Let

$$\alpha_i = \min_{x \in A} g_i(x) = g_i(x'), \qquad \beta_i = \max_{x \in A} g_i(x) = g_i(x'').$$

It is easy to prove that

$$(32)\quad E\{[g_i(X) - \lambda_i]^2 \mid F\} \leqq (\beta_i - \lambda_i)(\lambda_i - \alpha_i).$$

Thus

$$(33)\quad R(\bar{f}, F) \leqq \sum_{i=1}^{m} c_i\{[(1 - an)\lambda_i - b_i]^2 + na^2(\beta_i - \lambda_i)(\lambda_i - \alpha_i)\}.$$

Putting

$$a = \frac{1}{n + \sqrt{n}}, \qquad b_i = \frac{s_i}{\sqrt{n} + 1},$$

we obtain

$$(34)\quad R(f^0, F) \leqq \frac{1}{4(\sqrt{n} + 1)^2} \sum_{i=1}^{m} c_i(\beta_i - \alpha_i)^2 = c.$$

Observe that if a distribution $\bar{F}$ of the random variable X is defined by

$$(35)\quad \begin{aligned} P(X = x') &= 1 - p, \\ P(X = x'') &= p. \end{aligned}$$

Then $\lambda_i = \alpha_i + (\beta_i - \alpha_i)p$, and equality obtains in (32); i.e.

$$(36)\quad R(f^0, \bar{F}) = c.$$

The distribution F depends on the parameter p. Since (34) and (36) hold, it is sufficient to show (as in Sec. 3) that there exists a distribution G of p for which (30) is Bayes—that is, a distribution G such that (30) minimizes the expected risk

$$\begin{aligned} r(f, G) &= E[R(f, \bar{F})] = \sum_{i=1}^{m} c_i\, E\{E[(\lambda_i - f_i)^2 \mid \bar{F}]\} \\ &= \sum_{i=1}^{m} c_i E\{E\{[\alpha_i + (\beta_i - \alpha_i)p - f_i]^2 \mid p\}\}. \end{aligned}$$

It is easy to verify that this happens for the distribution $G^0(p)$ defined by equation

$$(37)\qquad dG^0(p) = C(pq)^{(\sqrt{n}/2)-1}\,dp \qquad (q = 1 - p).$$

This completes the proof.

6. In this paper, we have used the loss $L = \sum_{i=1}^{m} c_i(f_i - \omega_i)^2$. This loss has been extensively investigated ([2], [4], [5], [6]). For many special problems, other loss functions might be used, for example,

$$L = \sum_{i=1}^{m} c_i \,|\, f_i - \omega_i \,|,$$

about which little is known at present.

Problems considered in this paper were suggested to me by L. J. Savage and H. Steinhaus. I am indebted to J. A. Dubay and L. J. Savage for help and suggestions made during the preparation of this paper.

REFERENCES

[1] D Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, Inc , New York, 1954.
[2] M. A. Girshick and L J. Savage, "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 53–73.
[3] P. R. Halmos, "The theory of unbiased estimation," *Ann. Math Stat.*, Vol. 17 (1946), pp. 34–43.
[4] J. L. Hodges, Jr., and E. L Lehmann, "Some problems in minimax point estimation," *Ann. Math. Stat.*, Vol 21 (1950), pp. 182–197.
[5] H. Steinhaus, "The problem of estimation," *Ann. Math Stat* , Vol. 28 (1957), pp. 633–648.
[6] S. Trybuła, "On minimaksowej estymacji parametrów w rozkładzie wielomianowym," to be published.
[7] A. Wald, "Contributions to the theory of statistical estimation and testing hypothesis," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 299–325.
[8] A. Wald, *Statistical Decision Functions*, John Wiley and Sons, Inc., New York, 1950.

A THEOREM ON FACTORIAL MOMENTS AND ITS APPLICATIONS

By P. V. Krishna Iyer

Defense Science Laboratory, New Delhi

0. Summary. The theorem that the sth factorial moment for the sum of N events is $s!$ times the sum of the expectations for any s of the events occurring simultaneously has been proved by induction. The applications of this result in obtaining easily the moments of a number of distributions arising from a sequence of observations belonging to two continuous populations and other cases have been demonstrated.

1. Introduction. A number of distributions arising from a sequence of n observations belonging to a binomial population have been considered by the writer [3], [4] in some of his earlier publications. The moments of these distributions were obtained by using the theorem that the sth factorial moment is equal to $s!$ times the expectation for s of the characters considered in the distribution. Thus for a sequence of observations consisting of A's and B's with the probabilities p and q respectively, the sth factorial moment for the distribution of the total number of AB and BA joins between successive observations is the expectation for s joins like AB and BA in the sequence. It can be seen that there are s different ways of obtaining s joins. They are:

(1) From $(s + 1)$ consecutive observations.

(2) From two sets of l_1 and l_2 consecutive observations such that $l_1 + l_2 - 2$ is equal to s.

(3) From three sets of l_1, l_2 and l_3 consecutive observations such that

$$l_1 + l_2 + l_3 - 3$$

is equal to s.

(4) From k sets of $l_1, l_2, \cdots, l_k$ consecutive observations subject to the condition

$$\sum_1^k l_r - k = s,$$

where k takes values 1 to s.

The sum of the expectations for 1, 2, 3, $\cdots$, s is equal to $1/s!$ (the sth factorial moment for the distribution of the total number of AB and BA joins of the sequence).

The theorem as it stands appears to be applicable only for the distributions arising from a binomial sequence consisting of A's and B's with fixed probabilities p and q. We shall show in this paper that this result can be applied for distributions arising from two samples belonging to populations with cumu-

Received December 19, 1956; revised May 29, 1957.

lative distribution functions F and G. Before discussing this aspect, we shall first give a rigorous proof of the theorem and then show how it can be applied for the case of continuous distributions. The use of the result for distributions arising from Markoff chain is also illustrated.

2. Statement of theorem and proof.

THEOREM. *The* sth *factorial moment about the origin of any statistic* X *which is the sum of* N *events, dependent or independent, is equal to* $s!$ *times the sum of the expectations for any* s *of the events occurring together.*

PROOF. Let the events be denoted by $x_1, x_2, \cdots, x_N$. As in the case of binomial distribution, assume that the x's take value 1 if the event occurs and zero otherwise. Define

$$X = \sum_1^N x_r,$$

$$E(X) = E(\sum x_r) = \sum E(x_r) \tag{1}$$

= the sum of the expectations of the different events

= the sum of the probabilities for the events to occur.

$$E(X^2) = E(\sum x_r)^2 = E(\sum x_r^2) + 2E(\sum x_r x_s), \qquad s > r.$$

Now

$$E(\sum x_r^2) = E(X);$$

hence

$$E(X^2) = E(X) + 2E(\sum x_r x_s)$$

or

$$E\{X(X-1)\} = 2\sum E(x_r x_s) \tag{2}$$

= 2 (sum of the expectations for any two of the events)

= 2 (sum of the probabilities for any two of the events to occur together).

$$E(X^3) = E(\sum x_r)^3 = E(\sum x_r^3) + 3E\sum (x_r^2 x_s) + 3E(\sum x_r x_s^2) + 6E(\sum x_r x_s x_t), \qquad t > s > r$$

since

$$E(\sum x_r^3) = E(X),$$

$$E(x_r^2 x_s) = E(x_r x_s),$$

$$E(x_r x_s^2) = E(x_r x_s),$$

$$E(X^3) = E(X) + 6E(\sum x_r x_s) + 3!E(\sum x_r x_s x_t).$$

Substituting the value of $E(\sum x_r x_s)$ from (2), we get

$$E(X^3) = E(X) + 3E\{X(X-1)\} + 3!E(\sum x_r x_s x_t)$$

or

$$(3) \quad E\{X(X-1)(X-2)\} = 3!\sum E(x_r x_s x_t)$$

= 3!(sum of the expectations for any three of the events)

= 3!(sum of the probabilities for any three of the events to occur together).

Thus the theorem holds good for $s = 1$ to 3.

It may be noted that the results given above hold good even without taking the expectation of both sides because the x's take values 1 or 0 only.

We shall now establish the general relation by induction. For this we show that if

$$(4) \qquad X^{[s]} = s!\,(\sum x_{t_1} x_{t_2} \cdots x_{t_s})$$

holds good for any value of s, it is true for $(s+1)$ also.

Multiplying both sides of (4) by X we get

$$\begin{aligned}[X^{[s]}X] &= s!\,(\sum x_{t_1} x_{t_2} \cdots x_{t_s})(\sum x_r) \\ &= (s+1)!\,(\sum x_{t_1} x_{t_2} \cdots x_{t_s} x_{t_{s+1}}) \\ &+ s!\,[\sum x_{t_1}^2 x_{t_2} \cdots x_{t_s} \\ &+ \sum x_{t_1} x_{t_2}^2 \cdots x_{t_s} + \cdots + \sum x_{t_1} x_{t_2} \cdots x_{t_s}^2] \\ &= (s+1)!\sum x_{t_1} x_{t_2} \cdots x_{t_{s+1}}) \\ &+ s!\,(\sum x_{t_1} x_{t_2} \cdots x_{t_s}).\end{aligned}$$

Substituting for $\sum x_{t_1} x_{t_2} \cdots x_{t_s}$ from (4), (5) reduces to

$$(6) \qquad X^{[s]}(X - s) = X^{[s+1]} = (s+1)!\,(\sum x_{t_1} x_{t_2} \cdots x_{ts+1}).$$

Taking the expectation of both sides

$$(7) \qquad E\{X^{[s+1]}\} = (s+1)!\sum E(x_{t_1} x_{t_2} \cdots x_{ts+1}).$$

Hence the theorem.

3. Applications. We shall now examine how the above result can be applied for obtaining easily the moments of a number of distributions including those arising from a simple Markoff chain. Some of the distributions considered here have been discussed by Wald and Wolfowitz [7], Mood [6], Mann & Whitney [5], and others.

(1) *Binomial distribution.* It is obvious that the rth factorial moment for the distribution of x, the number of successes out of n trials is given by

(8) $$\frac{\mu'_{[r]}}{r!} = \binom{n}{r} p^r,$$

where p is the probability for a success.

(2) *Hypergeometric distribution.* This can be deduced from the above by substituting

$$p^r = \frac{(N - M)^{[r]}}{N^{[r]}},$$

where N and M have the usual significance This follows from the fact that the probability p for the *1st*, *2nd*, *3rd*, $\cdots$ successes are

$$\frac{N - M}{N}, \quad \frac{N - M - 1}{N - 1}, \quad \frac{N - M - 2}{N - 2}, \cdots$$

(3) *Distribution of the number of AB joins between successive observations of a binomial sequence.* We first note that r AB joins can be formed from only r sets of two consecutive observations each and therefore

(9) $$\frac{\mu'_{[r]}}{r!} = \binom{n - r}{r} p^r q^r.$$

This can be seen from the fact that the probabilities for an AB join is pq and that there are $\binom{n-r}{r}$ ways of obtaining them from n observations in a sequence.

(4) *Distribution of AB joins for binomial sequence of n_1A's and n_2B's.* As in the case of hypergeometric series, we substitute

$$p^r q^s = \frac{n_1^{[r]} n_2^{[s]}}{(n_1 + n_2)^{[r+s]}}$$

in the results given in (3) above. Thus

(10) $$\mu'_{[r]} = \frac{(n_1 + n_2 - r)^{[r]} n_1^{[r]} n_2^{[r]}}{(n_1 + n_2)^{[2r]}}.$$

(5) *Distribution of AB and BA joins between consecutive observations of a binomial sequence.* Taking for simplicity the third factorial moment, we note that three joins can be obtained from (i) four consecutive observations $ABAB$ or $BABA$, (ii) two sets, one of two and the other of three consecutive observations like $AB - ABA$; $BA - ABA$; $AB - BAB$ and $BA - BAB$, (iii) three sets of each of two consecutive observations AB or BA. The sum of the expectations for the above three ways of obtaining three joins is

(11) $$\frac{\mu'_{[3]}}{3!} = 2(n - 3)p^2q^2 + 8\binom{n-3}{2} p^2q^2 + \binom{n-3}{3} 8p^3q^3.$$

(6) *A sequence formed by pooling two samples A and B belonging to F.* Let two samples A and B of sizes n_1 and n_2 be drawn from a population where ...

distribution function is $F(x)$, $F(x)$ being continuous and x taking values from $-\infty$ to $+\infty$. By pooling together A and B and arranging them in ascending or descending order we obtain a sequence of A's and B's as in (4) considered above. Hence the moments of any distribution arising from this sequence can be obtained from the corresponding ones for the binomial sequence by substituting

$$p^r q^s = \frac{n_1^{[r]} n_2^{[s]}}{(n_1 + n_2)^{[r+s]}}.$$

(7) *Same as* (6), $F \neq G$. The calculation of the moments for some of the distributions discussed above is more complicated when $F \neq G$. We shall show below how the present theorem enables us to obtain the moments of these distributions also.

(a) *Number of observations of sample A to the left of the* rth *value of the combined ordered sequence of A's and B's.*

$\frac{\mu'_{[s]}}{s!}$ = [Number of ways of selecting A's from $(r-1)$ observations]
× [Probability that s out of the $(r-1)$ values to the left of the rth observation belong to A]

Assuming $n_1F(\alpha) + n_2G(\alpha) = r$, the probability that amongst the $(r-1)$ values to the left of the rth observation there are s A's is

$$\frac{n_1F(\alpha)}{n_1 F(\alpha) + n_2 G(\alpha)} \frac{(n_1-1)F(\alpha)}{(n_1-1)F(\alpha) + n_2 G(\alpha)} \frac{(n_1-2)F(\alpha)}{(n_1-2)F(\alpha) + n_2 G(\alpha)} \cdots s \text{ terms.}$$

Using the relation between $F(\alpha)$, $G(\alpha)$ and r we get

$$(12) \qquad \frac{\mu'_{[s]}}{s!} = \binom{r-1}{s} \frac{n_1^{[s]}[F(\alpha)]^s}{r[r-F(\alpha)][r-2F(\alpha)] \cdots [r-(s-1)F(\alpha)]}.$$

(b) *Number of AB and BA joins between successive observations.* As the higher moments are complicated we shall be content to obtain the second moment.

$\frac{\mu'_{[2]}}{2!}$ = the sum of the expectations for two joins from (i) three consecutive observations and (ii) two sets each of two consecutive observations.

Expectation for two joins from three consecutive observations x_1, x_2, and x_3 is given by

$$(13) \qquad \begin{aligned} &n_1(n_1-1)n_2 \int_{-\infty}^{+\infty} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} \{1 - F(x_3) + F(x_1)\}^{n_1-2} \\ &\quad \cdot \{1 - G(x_3) + G(x_1)\}^{n_2-1}\, dF(x_1)\, dG(x_2)\, dF(x_3) \\ &\quad + n_1n_2(n_2-1) \int_{-\infty}^{+\infty} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} \{1 - G(x_3) + G(x_1)\}^{n_2-2} \\ &\quad \cdot \{1 - F(x_3) + F(x_1)\}^{n_1-1}\, dG(x_1)\, dF(x_2)\, dG(x_3), \qquad x_1 < x_2 < x_3. \end{aligned}$$

Expectation for four joins from two sets of two consecutive observations each is equal to

$$(14)\quad \begin{aligned} n_1^{[2]}n_2^{[2]}\Bigg[\int_{-\infty}^{+\infty}\int_{-\infty}^{x_4}\int_{-\infty}^{x_3}\int_{-\infty}^{x_2} & A\, dF(x_1)\, dG(x_2)\, dF(x_3)\, dG(x_4) \\ & + \int_{-\infty}^{+\infty}\int_{-\infty}^{x_4}\int_{-\infty}^{x_3}\int_{-\infty}^{x_2} A\, dF(x_1)\, dG(x_2)\, dG(x_3)\, dF(x_4) \\ & + \int_{-\infty}^{+\infty}\int_{-\infty}^{x_4}\int_{-\infty}^{x_3}\int_{-\infty}^{x_2} A\, dG(x_1)\, dF(x_2)\, dF(x_3)\, dG(x_4) \\ & + \int_{-\infty}^{+\infty}\int_{-\infty}^{x_4}\int_{-\infty}^{x_3}\int_{-\infty}^{x_2} A\, dG(x_1)\, dF(x_2)\, dG(x_3)\, dF(x_4), \end{aligned}$$

where

$$\begin{aligned} A = \{1 - F(x_2) + F(x_1) - F(x_4) + F(x_3)\}^{n_1-2} & \\ \times \{1 - G(x_2) + G(x_1) - G(x_4) + G(x_3)\}^{n_2-2}, & \\ x_1 < x_2 < x_3 < x_4. & \end{aligned}$$

From the above it follows that

$$(15)\qquad \frac{\mu'_{[2]}}{2!} = (13) + (14).$$

When $F = G$, this reduces to the expression known.

(8) *Mann and Whitney's T-statistic.* In this case the expression for the second factorial moment reduces to the simple form

$$(16)\quad \begin{aligned} \mu'_{[2]} = 2n_1n_2 \int_{-\infty}^{+\infty}\int_{-\infty}^{x_3}\int_{-\infty}^{x_2} & [(n_1 - 1)f(x_1)f(x_2)g(x_3) + (n_2 - 1)f(x_1)g(x_2)g(x_3)] \\ & \cdot dx_1\, dx_2\, dx_3 + n_1^{(2)}n_2^{(2)}\left[\int_{-\infty}^{+\infty}\int_{-\infty}^{x_2} f(x_1)g(x_2)\, dx_1\, dx_2\right]^2, \end{aligned}$$

where $f(x)$ and $g(x)$ are the density functions for F and G.

(9) *AB joins between successive observations for a simple Markoff chain.* Let the matrix of probabilities for a simple Markoff chain be

$$\begin{bmatrix} p_1 & p_2 \\ q_1 & q_2 \end{bmatrix}.$$

Taking the probability that the first observation is A or B as P and Q respectively, the probabilities $P_r(A)$ and $Q_r(B)$ that the rth observation is A or B are given by

$$(17)\qquad \begin{aligned} P_r(A) &= \frac{p_2}{1 - \delta} + \frac{Pq_1 - Qp_2}{1 - \delta}\,\delta^{r-1}, \\ Q_r(B) &= 1 - P_r(A), \end{aligned}$$

where

$$\delta = p_1 - p_2 \quad \text{and} \quad p_1 > p_2 .$$

When the first observation is B, the conditional probability for the rth observation to be A reduces to

$$(18) \qquad P_r(A \mid 1B) = \frac{p_2}{1-\delta}(1 - \delta^{r-1}).$$

This is the same as given by Bartlett:

In the case of the Markoff chain, unlike the previous cases discussed earlier the probability of an AB join depends on the position of A in the sequence, and the expectation for two AB joins is given by

$$\begin{aligned} &q_1^2[P_1(A)\{P_3(A \mid 2B) + P_4(A \mid 2B) + P_5(A \mid 2B) + \cdots + P_{n-1}(A \mid 2B)\} \\ &+ P_2(A)\{P_4(A \mid 3B) + P_5(A \mid 3B) + P_6(A \mid 3B) + \cdots + P_{n-1}(A \mid 3B)\} \\ (19) \quad &+ P_3(A)\{P_5(A \mid 4B) + P_6(A \mid 4B) + \cdots + P_{n-1}(A \mid 4B)\} \\ &\cdots \\ &+ P_{n-2}(A)\{P_{n-1}(A \mid n-2, B)\}], \end{aligned}$$

where $P_r(A \mid B)$ is the conditional probability that the rth observation is A given that the sth observation is B when $r > s$. Summing up the above series after substituting for P's from (16) and (17), we get

$$\begin{aligned} \frac{\mu'_{[2]}}{2!} = \Bigg[&\frac{(n-2)(n-3)\alpha}{2} - \frac{\alpha\delta}{1-\delta}\left\{(n-3) - \frac{\delta(1-\delta^{n-3})}{1-\delta}\right\} \\ (20) \qquad &- \frac{\beta\delta}{1-\delta}\left\{\frac{1-\delta^{n-3}}{1-\delta} - (n-3)\delta^{n-3}\right\} \\ &+ \beta\left\{\frac{n-3}{1-\delta} - \frac{\delta(1-\delta^{n-3})}{(1-\delta)^2}\right\}\Bigg] \frac{p_2 q_1^2}{1-\delta} \end{aligned}$$

where

$$\alpha = \frac{p_2}{1-\delta} \quad \text{and} \quad \beta = \frac{Pq_1 - Qp_2}{1-\delta}.$$

It may be added that the result given in this paper can be used for deriving the moments of many other distributions of similar kind.

REFERENCES

1. M. S. BARTLETT, *An Introduction to Stochastic Processes*, Cambridge University Press, London, 1955, p. 29.
2. W. J. DIXON, "A criterion for testing the hypothesis that two samples are from the same population", *Ann. Math. Stat.*, Vol. 11 (1940), p. 199.
3. P. V. KRISHNA IYER, "The theory of probability distributions of points on a line", *J. Indian. Soc. Agri. Stat.*, Vol. 1 (1948), p. 173.

4 P. V. KRISHNA IYER, "The first and second moments of some probability distribution arising from points on a lattice and their applications", *Biometrika*, Vol. 36 (1949), p. 135
5 H. B. MANN AND D. R. WHITNEY, "On a test whether one of two random variables is stochastically larger than the other," *Ann. Math Stat.*, Vol. 18 (1947), p 50.
6 A. M. MOOD, "On the asymptotic efficiency of certain nonparametric two sample tests", *Ann. Math. Stat.*, Vol 25 (1954), p 514.
7. A. WALD AND J. WOLFOWITZ, "On a test whether two samples are from the same population," *Ann. Math Stat*, Vol 11 (1940), p 147

THE UNIQUENESS OF THE TRIANGULAR ASSOCIATION SCHEME

By W. S. Connor

National Bureau of Standards

1. Summary. Parameters for a class of partially balanced incomplete block designs with two associate classes are immediately implied by the triangular association scheme. This paper deals with the more difficult question of whether or not these parameters imply the triangular association scheme.

2. Introduction. A partially balanced incomplete block design with two associate classes [1] is said to be triangular [2], [3] if the number of treatments $v = n(n - 1)/2$ and the association scheme is an array of n rows and n columns with the following properties:

(a) The positions in the principal diagonal are blank.

(b) The $n(n - 1)/2$ positions above the principal diagonal are filled by the numbers $1, 2, \cdots, n(n - 1)/2$ corresponding to the treatments.

(c) The positions below the principal diagonal are filled so that the array is symmetrical about the principal diagonal.

(d) For any treatment i the first associates are exactly those treatments which lie in the same row and the same column as i.

The following relations clearly hold:

(1) The number of first associates of any treatment is $n_1 = 2n - 4$.

(2) With respect to any two treatments θ_1 and θ_2 which are first associates, the number of treatments which are first associates of both θ_1 and θ_2 is

$$p_{11}^1(\theta_1, \theta_2) = n - 2.$$

(3) With respect to any two treatments θ_3 and θ_4 which are second associates, the number of treatments which are first associates of both θ_3 and θ_4 is $p_{11}^2(\theta_3, \theta_4) = 4$.

We wish to examine the converse, i.e., whether or not relations (1), (2) and (3) imply (a), (b), (c), and (d). We shall give a proof for $n \geqq 9$ which shows that the converse is true. The cases with $n < 9$ will not be considered, although the author has found that it is true for several small values of n, and conjectures that it is true for the rest.

As background for this problem, it is interesting to recall what has been found for some other classes of partially balanced designs. In the analogous problem for the group divisible designs it is easy to show that the converse is true [4]. For the latin square designs the converse is true for a sufficiently large number of treatments, but is not always true, as has been shown by example [5].

The present problem is closely related to problems considered in [6] and [7].

Received July 8, 1957.

The arguments used here could be substituted for some of the arguments in those papers.

3. A characterization of the triangular association scheme. The proof will consist of showing that there exist sets of treatments which satisfy the following theorem.

THEOREM. *The triangular association scheme for $n(n-1)/2$ treatments exists if and only if there exist sets of treatments S_j, $j = 1, \cdots, n$, such that:*

(i) *Each S_j consists of $n - 1$ treatments.*

(ii) *Any treatment is in precisely two sets S_j.*

(iii) *Any two distinct sets S_i, S_j have exactly one treatment in common.*

Proof. Necessity follows from the observation that the n rows of treatments in the triangular association scheme are the n sets S_j.

Sufficiency follows from noting a correspondence between the rows and columns of the association scheme and the sets S_j. To display the correspondence, we denote the unique element common to sets S_i and S_j by $\alpha(i, j) = \alpha(j, i)$. Then the correspondence is as follows. We let set S_i correspond to the ith row and column, and put element $\alpha(i, j)$ in the ith row and jth column of the association scheme. Because $\alpha(i_1, j_1) = \alpha(i_2, j_2)$ implies that $i_1 = i_2$ and $j_1 = j_2$, the element $\alpha(i, j)$ occurs only in the ith row (column) and jth column (row). This fills up the association scheme as described in (a), (b) and (c). Further, if we let "belonging to the same set S_j" correspond to "being first associates", then (d) is satisfied.

4. The existence of sets S_j which satisfy the Theorem. In this section we shall show for $n \geqq 9$ that there exist sets S_j which satisfy the Theorem. The proof makes conspicuous use of the condition (3) that $p_{11}^2 = 4$. In fact, in constructing the proof, the author was attracted to the singular fact that this parameter does not depend on n.

Throughout the proof, we shall employ certain conventions. In citing a reason why something is or is not true, we often shall write "$p_{11}^1(\theta_1, \theta_2)$" or "$p_{11}^2(\theta_3, \theta_4)$," whereby we mean to refer to particular treatments θ_1, θ_2, θ_3, and θ_4. Also, we shall write "$(\theta_1, \theta_2) = 1$ (or 2)," meaning that treatments θ_1 and θ_2 are first (or second) associates.

In developing the proof, the author used a matrix in which the ith row and column correspond to the ith treatment, and the entry in the intersection of the ith row and jth column is 1 or 2, depending on whether treatments i and j are first or second associates. Though this matrix is not explicitly used below, it is implicit, and it is believed that the reader will find the use of this matrix helpful in following the proof.

We begin by proving a lemma which will be used repeatedly in the sequel.

LEMMA 1. *With respect to any two initial treatments θ_1 and θ_2 which are first associates, the $n - 3$, $(n \geqq 9)$ treatments which are first associates of θ_1 and second associates of θ_2 pairwise are first associates.*

Proof. For simplicity we shall replace θ_1 by 1 and θ_2 by 2. From (1) and (2)

it follows that there are $n - 2$ treatments which are first associates of both treatments 1 and 2, and $n - 3$ treatments which are first associates of treatment 1 and second associates of treatment 2. We shall refer to the treatments of the first set as treatments $3, \cdots, n$; and to those of the second set as treatments $n + 1, \cdots, 2n - 3$. These sets will be denoted respectively by

$$T_1 = T_1(3, \cdots, n)$$

and $T_2 = T_2(n + 1, \cdots, 2n - 3)$.

We first show that any treatment α in T_2 cannot have more than one second associate in T_2. We observe that $p_{11}^2(2, \alpha) = 4$, of which one such treatment is treatment 1. Thus, treatment α has at most three first associates in T_1. Because $p_{11}^1(1, \alpha) = n - 2$, treatment α has at least $n - 5$ first associates in T_2, and hence at most one second associate in T_2.

We now shall show that even this one second associate is impossible. Consider any two treatments α and β in T_2, and assume that $(\alpha, \beta) = 2$. We have established that treatment 1 and the $n - 5$ treatments other than α and β in T_2 are first associates of both α and β. But for $n \geqq 9$ the condition that $p_{11}^2(\alpha, \beta) = 4$ is violated, which shows that $(\alpha, \beta) = 1$. This completes the proof of Lemma 1.

Our next lemma shows the existence of sets S, which satisfy (i) and (ii) of the theorem.

LEMMA 2. *For $n \geqq 9$, any initial treatment θ is an element of exactly two sets of treatments S_1 and S_2 which are such that a set contains $n - 1$ treatments, the treatments in a set pairwise are first associates, and θ is the unique element common to S_1 and S_2.*

Proof. We begin by showing that Lemma 1 implies that there are $n - 4$ treatments in T_1 which pairwise are first associates. For this purpose, it is convenient to define sets $T_1' = T_1'(3, \cdots, n - 2)$ and

$$T_2' = T_2'(n + 2, \cdots, 2n - 3).$$

From Lemma 1 and the condition that $p_{11}^1(1, \alpha) = n - 2$ for every treatment α in T_2, it follows that every treatment in T_2 has two first associates and $n - 4$ second associates in T_1. Without essential loss of generality, let treatment $n + 1$ be a second associate of every treatment in T_1', and let $(n - 1, n + 1) = (n, n + 1) = 1$. Then by Lemma 1, letting $\theta_1 = 1$ and $\theta_2 = n + 1$, the treatments in T_1' pairwise are first associates.

We still have to determine how treatments $n - 1$ and n intersect the treatments in T_1', T_2' and each other. We shall show that $(n - 1, n) = 2$ and either we have Case 1: $(n - 1, \alpha) = 1$, $(n, \alpha) = 2$ for all treatments α in T_1' and $(n - 1, \beta) = 2$, $(n, \beta) = 1$ for all treatments β in T_2'; or we have Case 2: $(n - 1, \alpha) = 2$, $(n, \alpha) = 1$ for all α in T_1' and $(n - 1, \beta) = 1$, $(n, \beta) = 2$ for all β in T_2'.

Suppose that treatment $n - 1$ is a second associate of some treatment in T_2',

say treatment β. We shall show that we have Case 1. From Lemma 1 and $p_{11}^1(1, \beta)$ it follows that treatment β has two first associates and $n - 4$ second associates among the treatments of T_1' and treatments $n - 1$ and n. Therefore, treatment β has at least $n - 6$ second associates in T_1'. Without essential loss of generality, let these be treatments $3, \cdots, n - 4$. Applying Lemma 1 with $\theta_1 = 1$ and $\theta_2 = \beta$, it follows that these treatments are first associates of treatment $n - 1$.

Suppose that $(n - 3, n - 1) = 2$. Then it would be necessary that $p_{11}^2(n - 3, n - 1) = 4$. However, treatments $1, \cdots, n - 4$ are first associates of both treatments $n - 3$ and $n - 1$, violating $p_{11}^2(n - 3, n - 1) = 4$ for $n \geq 9$. Similarly, treatment $n - 2$ cannot be a second associate of treatment $n - 1$.

We have shown that if treatment $n - 1$ has a second associate in T_2', then it is a first associate of every treatment in T_1'. Further, the treatments in T_1' and treatments 2 and $n + 1$ satisfy the condition that $p_{11}^1(1, n - 1) = n - 2$, implying that treatment $n - 1$ is a second associate of treatment n and the treatments in T_2'.

By applying Lemma 1 with $\theta_1 = 1$ and $\theta_2 = n - 1$, it follows that $(n, \beta) = 1$ for all β in T_2'. Because treatment 2 and the treatments in T_2 satisfy $p_{11}^1(1, n)$, it follows that $(n, \alpha) = 2$ for all α in T_1'. This demonstrates Case 1.

If $(n - 1, \beta) \neq 2$ for any β in T_2', then $(n - 1, \beta) = 1$ for all β in T_2'. But treatment 2 and the treatments in T_2 satisfy $p_{11}^1(1, n - 1)$, and it follows that $(n - 1, \alpha) = 2$ for all α in T_1' and that $(n - 1, n) = 2$. We now apply Lemma 1 with $\theta_1 = 1$ and $\theta_2 = n - 1$ to show that $(n, \alpha) = 1$ for all α in T_1'. Because treatments $2, \cdots, n - 2, n + 1$ satisfy $p_{11}^1(1, n) = n - 2$, it follows that $(n, \beta) = 2$ for all β in T_2'. This establishes Case 2.

We now observe a set S_1 which contains treatments 1, 2, the treatments in T_1' and either treatment $(n - 1)$ or n. Also, a set S_2 which contains treatments 1, the treatments in T_2, and the one of treatments $n - 1$ and n which is not in S_1. These sets are such *that their elements pairwise are first associates. They* are the sets of Lemma 2.

To show that there are no other such sets, we shall consider the way in which the treatments in T_1' are associated *with the treatments in T_2'. Consider any* treatment α in T_1', and the condition $p_{11}^1(1, \alpha) = n - 2$. Treatment 2, the remaining $(n - 5)$ treatments in T_1', and either treatment $n - 1$ or n are $n - 3$ treatments which satisfy this condition. *Hence there is exactly one more such* treatment in T_2'. Similarly, any treatment β in T_2' has *exactly one first associate* in T_1'. It follows that no other set of $n - 1$ treatments exists such that its treatments pairwise are first associates. *This completes the proof of Lemma 2.*

The sets found in Lemma 2 obey (i) and (ii) *of the Theorem. To find the num*ber s of sets and to prove (iii), we observe that *each of s sets contains $n - 1$ elements*, so that there are $s(n - 1)$ *(not necessarily distinct) elements in the s sets.* But every treatment *occurs in exactly two sets, so that $s(n - 1) = 2r = n(n - 1)$* or $s = n$. Thus *the number of pairs of sets is $n(n - 1)/2 = r$, and because every treatment occurs in exactly two sets, we have (iii).*

REFERENCES

[1] R. C. Bose and K. R. Nair, "Partially balanced incomplete block designs," *Sankhya*, Vol. 4 (1939), pp. 337–372.

[2] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–190.

[3] R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, *Tables of Partially Balanced Designs with Two Associate Classes*, Institute of Statistics of The Consolidated University of North Carolina Reprint Series No. 50, 1954, p. 43.

[4] R. C. Bose and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[5] Dale Marsh Mesner, *An Investigation of Certain Combinatorial Properties of Partially Balanced Incomplete Block Experimental Designs and Association Schemes, with a Detailed Study of Designs of Latin Square and Related Types*, unpublished thesis, Michigan State University, 1956.

[6] W. S. Connor, Jr., "On the structure of balanced incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 57–71.

[7] Marshall Hall, Jr., and W. S. Connor, "An embedding theorem for balanced incomplete block designs," *Canadian Journal of Mathematics*, Vol. 6 (1953), pp. 35–41.

THE LIMITING DISTRIBUTION OF BROWNIAN MOTION IN A BOUNDED REGION WITH INSTANTANEOUS RETURN[1]

By B. Sherman

Westinghouse Research Laboratories

1. Summary. A point executes Brownian motion in a bounded, connected, and open three dimensional region D. When it reaches the boundary Γ, at point α, it is instantaneously returned to D according to probability measure $\mu(\alpha)$ (we write $\mu(\alpha, A)$ for the measure of set A), and the Brownian motion is resumed. This is a Markov process and, subject to certain regularity conditions on Γ and $\mu(\alpha)$, we derive the limiting distribution of the process. Processes of this sort have been considered by Feller [1]; he has obtained the transition probabilities of such processes. He is concerned more generally with Markov processes with continuous sample functions on a linear interval; the return may be instantaneous or after a random period of time.

Let $p^0(t, \xi, A)$ be the probability that the point is in set A of D at time t when it is initially at point ξ of D, with the additional restriction that no boundary contacts have been made. It is known that

$$p^0(t, \xi, A) = \int_A u(t, \xi, x)\, dx, \tag{1}$$

where dx is the volume element about x and u is the solution of the equation

$$\tfrac{1}{2}\Delta u = u_t ,$$

subject to the conditions

$$u(t, \xi, \alpha) = 0, \qquad \alpha \in \Gamma, \qquad \lim_{t \to 0} \int_C u(t, \xi, x)\, dx = 1,$$

where C is any sphere of non-zero radius with center ξ which is entirely within D. We may write explicitly

$$u(t, \xi, x) = \sum_{k=1}^{\infty} v_k(\xi) v_k(x) e^{-\lambda_k t},$$

where λ_k is the kth eigenvalue and $v_k(x)$ the corresponding eigenfunction of the equation $\Delta u + 2\lambda u = 0$ subject to the boundary condition $u = 0$ on Γ. If $K(\xi, x)$ is the Green's function of $\Delta u = 0$ in D, then[2] ([2], and [3], page 273)

Received February 9, 1956; revised June 27, 1957.

[1] Part of this paper was prepared under the sponsorship of the Office of Naval Research and the Office of Ordnance Research, U.S. Army, at the University of California, Berkeley and Los Angeles. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] I am indebted to Professors M. Kac and R. J. Duffin for some helpful discussions on the relation of the fundamental solution of the heat equation to the Green's function of the Laplace equation.

(2) $$K(\xi, x) = \frac{1}{2}\int_0^\infty u(t, \xi, x)\, dt.$$

Let $\phi(t, \xi, \alpha)\, dt\, d\alpha$ be the probability the point is absorbed at surface element $d\alpha$ of Γ between t and $t + dt$ when it is initially at point ξ of D. Then ϕ is half the interior normal derivative of u at point α of Γ([3], page 273). When the point is initially at ξ the probability of ultimate absorption in set S of Γ is given by

(3) $$\pi(\xi, S) = \int_S \int_0^\infty \phi(t, \xi, \alpha)\, dt\, d\alpha.$$

We may define a discrete parameter Markov process with Γ as state space by taking as transition probability

(4) $$\pi(\alpha, S) = \int_D \pi(\xi, S)\mu(\alpha, d\xi).$$

This Markov process has a limiting distribution π which satisfies the equation

(5) $$\pi(S) = \int_\Gamma \pi(\alpha, S)\pi(d\alpha).$$

We define a measure of sets of D by

$$\lambda(A) = \int_\Gamma \mu(\alpha, A)\pi(d\alpha).$$

We may now write the density function for the limiting distribution. If $M(\xi)$ is the mean time of reaching the boundary when the point is initially at ξ,

$$M(\xi) = \int_\Gamma \int_0^\infty t\phi(t, \xi, \alpha)\, dt\, d\alpha$$

then the density function of the limiting distribution is

(6) $$\frac{2\int_D K(\xi, x)\lambda(d\xi)}{\int_D M(\xi)\lambda(d\xi)}.$$

If we are given a probability measure λ in D and the return is always according to λ, then it is clear that the limiting density of this process is also given by (6). If λ concentrates at a single point ξ we may drop the integrals in (6), and in particular we get

$$M(\xi) = 2\int_D K(\xi, x)\, dx.$$

We note that (6) is essentially the steady distribution of temperature in the following problem: D is a homogeneous heat conducting body whose boundary is kept at temperature 0 and in which there is a constant source of heat distributed according to λ.

Regarding the regularity conditions, we shall assume that Γ is made up of finitely many surfaces, each with a continuously turning tangent plane and that D has a Green's function ([4], page 262). We will assume there is a closed set B in D such that

$$\inf_{\alpha\epsilon\Gamma} \mu(\alpha, B) = \gamma > 0.$$

2. Origin of the problem. This problem had its origins in the ecological research of Professor Thomas Park of the University of Chicago. He has been investigating problems of population stability and inter-species competition of flour beetles. It was discovered, on statistical investigations suggested in part by Jerzy Neyman, that the distribution of the beetles in the container of flour was not uniform, with the density increasing toward the boundaries of the container. The problem arose as to whether the nonuniformity might be simply a consequence of the random motion of the beetles or whether it ought to be attributed to some inhomogeneity such as a temperature gradient in the flour. To check the plausibility of the idea that the nonuniformity might arise from random motion alone, we have set up a model which may have some relevance to the actual situation. The region D represents the volume of flour. We assume the independence of the motions of the beetles so that we may confine ourselves to the random motion of a single point This is a reasonable assumption if the density of the beetles is low For the random motion we take Brownian motion, this is appropriate if we want path continuity and spatial homogeneity Finally, we must introduce some mechanism of return from the boundary, we use the device of instantaneous return If the return distribution is concentrated near the point of contact on the boundary then the device has some semblance of plausibility. More precisely we may suppose $\mu(\alpha, A(\alpha)) = 1$, where $A(\alpha)$ is that set of points of D whose distance from α is less than or equal to δ, a small positive number. Then if E is the subset of points of D whose distance from Γ is in excess of δ we have $\lambda(E) = 0$. If we are prepared to accept the density of distribution in E, as given by (6), as a theoretical model for what is observed then we are faced with a contradiction. For the density is a harmonic function in E, by virtue of $\lambda(E) = 0$. Because of the minimum-maximum properties of such functions we cannot have increasing density from the central parts of E outward to the boundary of E since that would entail a minimum at an interior point of E.

3. Derivation of the limiting distribution. We sketch a proof that we have defined a process by the instantaneous return mechanism. This is equivalent to proving that finitely many contacts occur in a finite time with probability 1. By the assumption on μ it will happen infinitely often with probability 1 that the point is returned to B. Let $T(x)$ be the time to reach Γ, starting at x. If δ is positive Prob $(T(x) > \delta)$ is a continuous function of x which achieves a positive minimum on B. Thus of the times the point is returned to B it will happen infinitely often with probability 1 that the time to reach Γ is in excess of δ. This implies that infinitely many contacts in a finite time has probability 0.

Let $p(t, x, A)$ be the transition probability of the process, i.e., the probability the point is in A at time t when it is initially at x. Then we prove the limiting distribution p exists and

$$p(t, x, A) = p(A) + f(t, x, A), \tag{7}$$

where

$$|f(t, x, A)| < ae^{-kt}. \tag{8}$$

Here a and k are positive and independent of x and A. To simplify the notation we will make the following convention: if $f(x)$ is a function on D then we define a corresponding function $f(\alpha)$ on Γ by taking the integral, over D, of f with respect to the measure $\mu(\alpha)$. With this convention we may replace x by α in (7) and (8). We note that both $f(t, x, A)$ and $f(t, \alpha, A)$ are integrable with respect to t from 0 to ∞.

Proceeding with the proof we use the fact that $u(t, \xi, x)$ is strictly positive for all ξ and x in D and for positive t. Then the minimum $v(x, \delta, t)$, achieved by $u(T, \xi, x)$ subject to $\xi \varepsilon B$ and $t - \delta \leqq T \leqq t$, is also strictly positive, and it follows directly that for $t - \delta \leqq T \leqq t$,

$$p^0(T, \alpha, A) \geqq \gamma \int_A v(x, \delta, t)\, dx.$$

It is clear that

$$h(\delta) = \inf_{x \varepsilon D} \int_\Gamma \int_0^\delta \phi(t, x, \alpha)\, dt\, d\alpha$$

satisfies $0 < h(\delta) < 1$ for all $\delta > 0$. Let $p^1(t, \xi, A)$ be the probability the point, initially at ξ, is in A at time t having made exactly one boundary contact. Then if $t > \delta$,

$$\begin{aligned} p^1(t, \xi, A) &= \int_\Gamma \int_0^t \phi(\tau, \xi, \alpha) p^0(t - \tau, \alpha, A)\, d\tau\, d\alpha \\ &\geqq \int_\Gamma \int_0^\delta \phi(\tau, \xi, \alpha) p^0(t - \tau, \alpha, A)\, d\tau\, d\alpha \\ &\geqq \int_\Gamma \int_0^\delta \phi(\tau, \xi, \alpha)\, d\tau\, d\alpha \cdot \gamma \int_A v(x, \delta, t)\, dx \\ &\geqq h(\delta)\gamma \int_A v(x, \delta, t)\, dx. \end{aligned}$$

We follow now the proof of a similar theorem given by Doob ([5], page 197). If $m(t, A)$ and $M(t, A)$ are respectively the infimum and supremum of $p(t, \xi, A)$ as ξ varies over D, then $M(t, A) \geqq m(t, A)$ and by the Chapman-Kolmogorov equation it can be seen that $M(t, A)$ is non-increasing and $m(t, A)$ is non-decreasing. For fixed t_0, ξ_0, x_0 define the set function

$$\psi(A) = p(t_0, \xi_0, A) - p(t_0, x_0, A).$$

There is a set A^+ on which ψ is maximum, such that $\psi(A) \geqq 0$ for any subset A of A^+, and such that $\psi(A) \leqq 0$ for any subset A of $A^- = D - A^+$. We have, *assuming* δ *such that* $0 < \delta < t_0$,

$$\begin{aligned}\psi(A^+) &= 1 - p(t_0, \xi_0, A^-) - p(t_0, x_0, A^+)\\ &\leqq 1 - p^1(t_0, \xi_0, A^-) - p^1(t_0, x_0, A^+)\\ &\leqq 1 - h(\delta)\gamma \int_D v(x, \delta, t_0)\, dx = c < 1.\end{aligned}$$

Following now a line of argument analogous to Doob's we have

$$M(t, A) - m(t, A) \leqq c^{(t/t_0)-1},$$

from which it follows that $M(t, A)$ and $m(t, A)$ have a common limit $p(A)$ and that

$$|p(t, x, A) - p(A)| \leqq M(t, A) - m(t, A) \leqq c^{(t/t_0)-1}.$$

Thus (7) and (8) are established, with $a = c^{-1}$ and $k = 1/t_0 \log c^{-1}$.

Before deriving (6) we have to establish the existence of the limiting distribution π of the boundary process. To this end we prove the lemma

(9) $$\zeta = \sup_{S \epsilon \Gamma} (\max_{x \epsilon B} \pi(x, S) - \min_{x \epsilon B} \pi(x, S)) < 1.$$

We note that $\pi(x, S)$ is, for fixed S, a harmonic function of x([3], page 273). If (9) is not true there will be a sequence of sets S_k such that

(10) $$\max_{x \epsilon B} \pi(x, S_k) \to 1, \qquad \min_{x \epsilon B} \pi(x, S_k) \to 0.$$

Since $\pi(x, S_k)$ is a sequence of harmonic functions with $0 \leqq \pi(x, S_k) \leqq 1$, we may extract a subsequence which converge to a harmonic function $f(x)$ uniformly on any compact subset of D([4], page 249). Without change of notation we suppose this done. However (10) implies that $f(x)$ achieves the values 1 and 0 on B, which contradicts the fact that $f(x)$ is harmonic in D and $0 \leqq f(x) \leqq 1$

To prove the existence of the limiting distribution of the boundary process we again follow the lines of Doob's proof. For fixed α and β we define the set function

$$\psi(S) = \pi(\alpha, S) - \pi(\beta, S).$$

Associated with ψ are the sets S^+ and S^-, and we have

$$\begin{aligned}\psi(S^+) &= 1 - (\pi(\alpha, S^-) + \pi(\beta, S^+))\\ &\leqq 1 - \left(\int_B \pi(x, S^-)\mu(\alpha, dx) + \int_B \pi(x, S^+)\mu(\beta, dx)\right)\\ &\leqq 1 - \gamma\, (\min_{x \epsilon B} \pi(x, S^-) + \min_{x \epsilon B} \pi(x, S^+))\\ &= 1 - \gamma(1 - \max_{x \epsilon B} \pi(x, S^+) + \min_{x \epsilon B} \pi(x, S^+))\\ &\leqq 1 - \gamma(1 - \zeta).\end{aligned}$$

If now we introduce $m^{(n)}(S)$ and $M^{(n)}(S)$, the infimum and supremum of the n step transition probability $\pi^{(n)}(\alpha, S)$ as α ranges over Γ, then following Doob's proof

$$M^{(n)}(S) - m^{(n)}(S) \leqq (1 - \gamma(1 - \zeta))^n,$$

from which it follows that the limiting distribution exists and satisfies (5).

We are now in position to derive (6). We have

$$p(t, \xi, A) = p^0(t, \xi, A) + \int_\Gamma \int_0^t \phi(\tau, \xi, \alpha)p(t - \tau, \alpha, A)\, d\tau\, d\alpha.$$

Introducing (7) and integrating with respect to t we get, after some reductions,

$$\begin{aligned}\int_0^T p^0(t, \xi, A)\, dt = p(A)\Big[T\Big(1 - \int_0^T \int_\Gamma \phi(\tau, \xi, \alpha)\, d\alpha\, d\tau\Big) \\ + \int_0^T \int_\Gamma \tau\phi(\tau, \xi, \alpha)\, d\alpha\, d\tau\Big] + I(T, \xi, A),\end{aligned} \tag{11}$$

where

$$I(T, \xi, A) = \int_0^T f(t, \xi, A)\, dt - \int_0^T \int_\Gamma \int_0^t \phi(\tau, \xi, \alpha)f(t - \tau, \alpha, A)\, d\tau\, d\alpha\, dt. \tag{12}$$

The second term in the bracket on the right of (11) tends to $M(\xi)$ as $T \to \infty$, and we show that the first term tends to 0. This term can be written

$$T \text{ Prob } (x(t) \,\varepsilon\, D, 0 < t \leqq T \mid x(0) = \xi). \tag{13}$$

Let the coordinates of point x be x_1, x_2, x_3 and suppose D is contained between the planes $x_1 = a$ and $x_1 = -a$. Then (13) tends to 0 if the expression

$$T \text{ Prob } (-a < x_1(t) < a, 0 < t \leqq T \mid x(0) = \xi) \tag{14}$$

tends to 0. We may write (14) explicitly

$$T \sum_{n=1}^{\infty} \frac{4}{(2n + 1)\pi} \sin \frac{(2n + 1)\pi}{2a} (\xi + a) \exp\left(-\frac{(2n + 1)^2\pi^2 T}{8a^2}\right),$$

which is less than

$$T \sum_{n=1}^{\infty} \frac{4}{(2n + 1)\pi} \exp\left(-\frac{(2n + 1)^2\pi^2 T}{8a^2}\right), \tag{15}$$

and it is easily proved that (15) tends to 0. Letting $T \to \infty$ in (11) we get

$$\int_0^\infty p^0(t, \xi, A)\, dt = p(A)M(\xi) + I(\infty, \xi, A). \tag{16}$$

Referring to (12) and (3) we may write, on introducing the variables $\tau' = \tau$ and $t' = t - \tau$,

$$I(\infty, \xi, A) = \int_0^\infty f(t, \xi, A)\, dt - \int_\Gamma \left(\int_0^\infty \phi(\tau', \xi, \alpha)\, d\tau' \right) \left(\int_0^\infty f(t', \alpha, A)\, dt' \right) d\alpha$$
$$= \int_0^\infty f(t, \xi, A)\, dt - \int_\Gamma \left(\int_0^\infty f(t, \alpha, A)\, dt \right) \pi(\xi, d\alpha).$$

The integration of the right side with respect to measure λ is equivalent to consecutive integrations with respect to $\mu(\beta)$ and π. The first integration gives, using (4),

$$\int_0^\infty f(t, \beta, A)\, dt - \int_\Gamma \left(\int_0^\infty f(t, \alpha, A)\, dt \right) \pi(\beta, d\alpha);$$

and the second, using (5), gives the value 0. Thus integrating on both sides of (16) with respect to measure λ we get

$$\int_D \int_0^\infty p^0(t, \xi, A)\, dt\, \lambda(d\xi) = p(A) \int_D M(\xi)\lambda(d\xi).$$

This equation, together with (1) and (2), implies that (6) is the density function of the limiting distribution p.

REFERENCES

[1] W. Feller, "Diffusion processes in one dimension," *Trans. Amer. Math. Soc.*, Vol. 77 (1954), pp. 1-31.
[2] G. A. Hunt, "On positive Green's functions," *Proc. Nat. Acad. Sci.*, Vol. 40 (1954), pp. 816-818.
[3] P. Lévy, *Processus Stochastiques et Mouvement Brownien*, Gauthier-Villars, Paris, 1948.
[4] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, Vol. 2, Julius Springer, Berlin, 1937.
[5] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, New York, 1953.

ON THE DISTRIBUTION OF A STATISTIC BASED ON ORDERED UNIFORM CHANCE VARIABLES

By Shanti S. Gupta[1] and Milton Sobel

Bell Telephone Laboratories

1. Summary. The exact distribution of a statistic based on the r smallest of n independent observations from a unit uniform distribution is derived. In life-testing terminology, this statistic includes as special cases (i) the sum of the r earliest failure times, (ii) the total observed life up to the rth failure, and (iii) the sum of all n failure times. The density, cumulative distribution function (c.d.f.) and first four moments of the general statistic are summarized in Sec. 2. Section 3 gives the derivation of the density and c.d.f. The moments are obtained from the moment generating function in Sec. 4. Asymptotic normality under certain conditions is proved in Sec. 5 and illustrations of the rapidity of approach to normality are given in Sec. 6.

2. Introduction and statement of results. We shall consider the statistic

$$T_{r,m}^{(n)} = t_1 + t_2 + \cdots + t_r + (m - r)t_r , \tag{2.1}$$

where $t_i = t_i^{(n)}$ is the ith smallest of n independent observations and m is greater than $r - 1$ but is not necessarily an integer. For $m = n$ this statistic can be interpreted as the total observed life in a life-testing experiment without replacement. When the underlying distribution of the unordered t's is exponential, i.e., $f(t) = (1/\theta)e^{-t/\theta}$, then it is known [3] that $2T_{r,n}^{(n)}/\theta$ is distributed as chi-square χ_{2r}^2 with $2r$ degrees of freedom.

Before stating further results let us introduce for $0 \leqq t \leqq m$ and non-negative integers p, q, n

$$A_{p,m}^{(q,n)}(t) = \frac{n}{p!}\left\{\binom{p}{0}\frac{t^{n-1}}{m^{q-p}} - \binom{p}{1}\frac{(t-1)^{n-1}}{(m-1)^{q-p}} + \binom{p}{2}\frac{(t-2)^{n-1}}{(m-2)^{q-p}} - \cdots\right\}, \tag{2.2}$$

where $m > p$, $n \geqq 1$ and the summation is continued as long as the arguments t, $t - 1$, $t - 2$, $\cdots$ are positive. It is understood that the binomial coefficient $\binom{p}{j} = 0$ for $p < j$ so that there are at most $(p + 1)$ terms in the above summation.

It is clear from (2.1) that $T_{n,n}^{(n)}$ is the sum of all the n observations. When the underlying distribution is unit uniform, then the density of $T_{n,n}^{(n)}$ is given on p. 246 of [2] by

$$f_{n,n}^{(n)}(t) = \frac{1}{(n-1)!}\left\{\binom{n}{0}t^{n-1} - \binom{n}{1}(t-1)^{n-1} + \cdots\right\} = A_{n,m}^{(n,n)}(t). \tag{2.3}$$

Received March 5, 1957, revised May 16, 1957.

[1] Now at the University of Alberta.

(We have removed the superscripts and subscripts from the chance variables and put them on f and on F below which are the symbols for the density and c d.f., respectively.)

Using the symmetry of the above density about $t = n/2$, we can replace t by $n - t$ in (2.3) obtaining

$$f^{(n)}_{n,n}(t) = \frac{n}{(n-1)!}\left\{\binom{n-1}{0}\frac{(n-t)^{n-1}}{n} - \binom{n-1}{1}\frac{(n-1-t)^{n-1}}{(n-1)} + \cdots\right\} = A^{(n,n)}_{n-1,n}(n-t), \tag{2.4}$$

where $0 \leqq t \leqq n$. The form (2.4) is more comparable with the results derived here. It is shown below that the density and c.d.f. of $T^{(n)}_{r,m}$ are given by the comparable results

$$f^{(n)}_{r,m}(t) = A^{(n,n)}_{r-1,m}(m-t) \tag{2.5}$$

and

$$F^{(n)}_{r,m}(t) = 1 - \frac{1}{(n+1)} A^{(n,n+1)}_{r-1,m}(m-t), \tag{2.6}$$

from which we get as special cases the densities and c.d.f.'s of (i) $T^{(n)}_{r,r}$, (ii) $T^{(n)}_{r,n}$ and (iii) $T^{(n)}_{n,n}$.

Barton and David [1] have derived another equivalent formula for the density $f^{(n)}_{r,r}(t)$, i.e., in the special case (i). Their result, with two typographical corrections taken into account, is

$$f^{(n)}_{r,r}(t) = \frac{n}{r!}\sum_{i=1}^{r}(-1)^{r-i} i^{r-n}\binom{r}{i}\left[\frac{i-t+|i-t|}{2}\right]^{n-1}. \tag{2.7}$$

The total life statistic arises as an optimum statistic under exponential distribution assumptions in [3]. In the present paper we give the distribution of this statistic when the exponential distribution assumption is replaced by the uniform distribution. Hence these results can be used to study the robustness of the tests based on the total life statistic. The results on asymptotic normality are also of interest in this connection since under the exponential assumption the distribution of $2T^{(n)}_{r,n}/\theta$ is that of χ^2_{2r}, which, for large r, also is close to that of a normal distribution. It is felt that the model of a uniform distribution from 0 to θ, $\theta > 0$ and unknown, and the results of this paper may prove to be useful in some life-testing problems.

3. Derivation of results. Let $u = t_1 + t_2 + \cdots + t_{r-1}$, $v = t_r$, $w = u/v$ and $y = T^{(n)}_{r,m} = u + (m - r + 1)v$, where $t_i = t_i^{(n)}$ is the ith smallest of n independent chance variables uniformly distributed from zero to one. The conditional distribution of w given v is exactly that of a sum of $r - 1$ independent uniform chance variables and is given by (2.3) with n replaced by $r - 1$. Hence the joint density of v and w is given by

$$
\begin{aligned}
g(w, v) &= \frac{n!}{(r-1)!\,(n-r)!} v^{r-1}(1-v)^{n-r} \frac{1}{(r-2)!} \\
&\quad \cdot \left\{ \binom{r-1}{0} w^{r-2} - \binom{r-1}{1} (w-1)^{r-2} + \cdots \right\},
\end{aligned} \tag{3.1}
$$

where $0 \leqq w \leqq r - 1$ and $0 < v \leqq 1$ and the joint density of u and v is given by

$$
\begin{aligned}
h(u, v) &= \frac{n!}{(r-2)!\,(r-1)!\,(n-r)!} \\
&\quad \cdot \left\{ \binom{r-1}{0} u^{r-2} - \binom{r-1}{1} (u-v)^{r-2} + \cdots \right\} (1-v)^{n-r},
\end{aligned} \tag{3.2}
$$

where $0 \leqq u \leqq (r-1)v$ and $v \leqq 1$.

If we now derive the density of y, then the full range of y from 0 to m is broken into r parts. For $0 \leqq y \leqq m - r + 1$, the density of y becomes

$$
\begin{aligned}
f_{r,m}^{(n)}(y) &= \frac{n!}{(r-2)!\,(r-1)!\,(n-r)!} \\
&\quad \cdot \int_{y/m}^{y/(m-r+1)} [y - (m - r + 1)v]^{r-2}(1-v)^{n-r}\, dv \\
&\quad - \binom{r-1}{1} \int_{y/m}^{y/(m-r+2)} [y - (m - r + 2)v]^{r-2}(1-v)^{n-r}\, dv \\
&\quad + \cdots + (-1)^{r-2} \binom{r-1}{r-2} \int_{y/m}^{y/(m-1)} [y - (m-1)v]^{r-2}(1-v)^{n-r}\, dv.
\end{aligned} \tag{3.3}
$$

Using the finite difference operators $\mathcal{E}$, Δ (with $\mathcal{E} = 1 + \Delta$),

$$
\begin{aligned}
f_{r,m}^{(n)}(y) &= \frac{r}{(r-2)!} \binom{n}{r} \sum_{\alpha=0}^{r-1} \binom{r-1}{\alpha} \\
&\quad \cdot (-\mathcal{E})^{\alpha} \left\{ \int_{y/m}^{y/(m-r+1+x)} [y - (m - r + 1 + x)v]^{r-2}(1-v)^{n-r}\, dv \right\},
\end{aligned} \tag{3.4}
$$

where $\mathcal{E}$ operates on x and it is understood that x is then to be set equal to 0. Using the relation between $\mathcal{E}$ and Δ,

$$
\begin{aligned}
f_{r,m}^{(n)}(y) &= \frac{r}{(r-2)!} \binom{n}{r} \\
&\quad \cdot \left[(-\Delta)^{r-1} \left\{ \int_{y/m}^{y/(m-r+1+x)} [y - (m - r + 1 + x)v]^{r-2}(1-v)^{n-r}\, dv \right\} \right]_{x=0}.
\end{aligned} \tag{3.5}
$$

If we now integrate by parts, the first term vanishes at the upper limit and also at the lower limit because of the operator Δ^{r-1}. After $r - 1$ such integrations we obtain

$$
f_{r,m}^{(n)}(y) = \frac{n}{(r-1)!} \left[\Delta^{r-1} \left\{ \frac{(m - r + 1 + x - y)^{n-1}}{(m - r + 1 + x)^{n-r+1}} \right\} \right]_{x=0}. \tag{3.6}
$$

Using $\Delta = E - 1$ we obtain

$$f_{r,m}^{(n)}(y)$$

(3.7)
$$= \frac{n}{(r-1)!}\left\{\binom{r-1}{0}\frac{(m-y)^{n-1}}{m^{n-r+1}} - \binom{r-1}{1}\frac{(m-1-y)^{n-1}}{(m-1)^{n-r+1}} + \cdots\right\}$$
$$= A_{r-1,m}^{(n,n)}(m-y),$$

where $0 \leqq y \leqq m - r + 1$ and $A_{r,m}^{(s,n)}$ is defined in (2.2).

We shall now show that the expression (3.7) gives the result for all y $(0 \leqq y \leqq m)$. For $m - r + i \leqq y \leqq m - r + 1 + i$ $(i = 1, 2, \cdots, r - 1)$ the only difference is that the first i upper limits of integration in (3.3) are all changed to unity. For the jth integral $(j = 1, 2, \cdots, i)$ we have to add to the complete set of r terms in (3.7) the quantity

$$(-1)^{j+1}\binom{r-1}{j-1}\frac{n!}{(r-2)!(r-1)!(n-r)!}$$

(3.8)
$$\cdot \int_{y/(m-r+1)}^{1} [y - (m - r + j)t]^{r-2}(1 - t)^{n-r}\, dt$$
$$= (-1)^{j+1}\frac{n}{(r-1)!}\binom{r-1}{j-1}\frac{(m-r+j-y)^{n-1}}{(m-r+j)^{n-r+1}}.$$

For each $j(1 \leqq j \leqq i)$ the quantity on the right in (3.8) cancels the jth term from the end of the complete expression with r terms in (3.7). Hence for

$$m - r + i \leqq y \leqq m - r + i + 1$$

the density is given by the first $r - i$ terms of (3.7) which are precisely those terms with positive arguments. This proves that the expression $A_{r-1,m}^{(n,n)}(m - y)$ of (3.7) gives the result for all y $(0 \leqq y \leqq m)$.

The c.d.f. $F_{r,m}^{(n)}(y)$ of y is easily obtained by integrating (3.7) between the limits 0 and y and is given by

(3.9)
$$F_{r,m}^{(n)}(y) = 1 - \frac{1}{(n-1)}A_{r-1,m}^{(n,n-1)}(m - y).$$

4. Moments of $y = T_{r,m}^{(n)}$. Using the expression for the density it can be shown that the moment generating function $M_t(y)$ of $y = T_{r,m}^{(n)}$ is given by

(4.1)
$$M_t(y) = \left[\frac{n}{(r-1)!}\sum_{\alpha=0}^{r-1}(-1)^{\alpha}\binom{r-1}{\alpha}\right.$$
$$\left.\cdot E^{\alpha}\left\{\frac{1}{(m-x)^{n-r+1}}\int_0^{m-x} e^{ty}(m-x-y)^{n-1}\,dy\right\}\right]_{x=0}$$

(4.2)
$$= \frac{n!}{(r-1)!}\sum_{j=r}^{\infty}\frac{(-t)^{j}}{(j-r)!}\Delta^{r-1}(x-m)^{j-r}\Big]_{x=0}.$$

$$g(w, v) = \frac{n!}{(r-1)!(n-r)!} v^{r-1}(1-v)^{n-r} \frac{1}{(r-2)!} \cdot \left\{ \binom{r-1}{0} w^{r-2} - \binom{r-1}{1} (w-1)^{r-2} + \cdots \right\}, \tag{3.1}$$

where $0 \leqq w \leqq r - 1$ and $0 < v \leqq 1$ and the joint density of u and v is given by

$$h(u, v) = \frac{n!}{(r-2)!(r-1)!(n-r)!} \cdot \left\{ \binom{r-1}{0} u^{r-2} - \binom{r-1}{1} (u-v)^{r-2} + \cdots \right\} (1-v)^{n-r}, \tag{3.2}$$

where $0 \leqq u \leqq (r-1)v$ and $v \leqq 1$.

If we now derive the density of y, then the full range of y from 0 to m is broken into r parts. For $0 \leqq y \leqq m - r + 1$, the density of y becomes

$$\begin{aligned} f_{r,m}^{(n)}(y) = {} & \frac{n!}{(r-2)!(r-1)!(n-r)!} \cdot \int_{y/m}^{y/(m-r+1)} [y - (m-r+1)v]^{r-2}(1-v)^{n-r}\, dv \\ & - \binom{r-1}{1} \int_{y/m}^{y/(m-r+2)} [y-(m-r+2)v]^{r-2}(1-v)^{n-r}\, dv \\ & + \cdots + (-1)^{r-2}\binom{r-1}{r-2} \int_{y/m}^{y/(m-1)} [y-(m-1)v]^{r-2}(1-v)^{n-r}\, dv. \end{aligned} \tag{3.3}$$

Using the finite difference operators $\mathcal{E}$, Δ (with $\mathcal{E} = 1 + \Delta$),

$$f_{r,m}^{(n)}(y) = \frac{r}{(r-2)!}\binom{n}{r} \sum_{\alpha=0}^{r-1} \binom{r-1}{\alpha} \cdot (-\mathcal{E})^{\alpha} \left\{ \int_{y/m}^{y/(m-r+1+x)} [y-(m-r+1+x)v]^{r-2}(1-v)^{n-r}\, dv \right\}. \tag{3.4}$$

where $\mathcal{E}$ operates on x and it is understood that x is then to be set equal t[illegible]
Using the relation between $\mathcal{E}$ and Δ,

$$f_{r,m}^{(n)}(y) = \frac{r}{(r-2)!}\binom{n}{r} \cdot \left[(-\Delta)^{r-1} \left\{ \int_{y/m}^{y/(m-r+1+x)} [y-(m-r+1+x)v]^{r-} \right. \right. \tag{3.5}$$

If we now integrate by parts, the first term vanishes [illegible]
at the lower limit because of the operator Δ^{r-1}. Aft[illegible]
obtain

$$f_{r,m}^{(n)}(y) = \frac{n}{(r-1)!}\left[\Delta^{r-1}\left\{\frac{(m}{} \right.\right. \tag{3.6}$$

Let

$$(5.1) \qquad y^* = \frac{y - E(y)}{\sigma(y)} \text{ and } y_v^* = \frac{y - E(y \mid v)}{\sigma(y \mid v)},$$

where $y = T_{r,m}^{(n)} = u_1 + u_2 + \cdots + u_{r-1} + (m - r + 1)v$.
The characteristic function of y^* is given by

$$\varphi_y^*(t) = \int_0^1 \int_0^v \cdots \int_0^v \exp\left\{it\left(\frac{y - E(y)}{\sigma(y)}\right)\right\}\left[\prod_{i=1}^{r-1} \frac{du_i}{v}\right] g(v)\, dv$$

$$(5.2) \qquad = \int_0^1 \int_0^v \cdots \int_0^v \exp\left\{it\, \frac{\sigma(y \mid v)}{\sigma(y)} \left[\frac{y - E(y \mid v)}{\sigma(y \mid v)}\right] + it\, \frac{\sigma(v)}{\sigma(y)} \left[\frac{E(y \mid v) - E(y)}{\sigma(v)}\right]\right\} \left[\prod_{i=1}^{r-1} \frac{du_i}{v}\right] g(v)\, dv,$$

where

$$(5.3) \qquad E(y \mid v) = (r - 1)\frac{v}{2} + (m - r + 1)v = (2m - r + 1)\frac{v}{2},$$

$$(5.4) \qquad \sigma(y \mid v) = v\sqrt{\frac{r-1}{12}}; \qquad \sigma(v) = \frac{1}{(n+1)}\sqrt{\frac{r(n - r + 1)}{n + 2}},$$

and

$$(5.5) \qquad g(v) = \frac{n!}{(r-1)!(n-r)!}\, v^{r-1}(1 - v)^{n-r}.$$

Letting

$$(5.6) \qquad t' = t\,\frac{\sigma(y \mid v)}{\sigma(y)}, \qquad t'' = t\,\frac{\sigma(v)}{\sigma(y)}\left[m - \left(\frac{r-1}{2}\right)\right],$$

and $x_i = u_i/v\ (i = 1, 2, \cdots, r - 1)$, we obtain

$$(5.7) \qquad \varphi_y^*(t) = \int_0^1 \left[\int_0^1 e^{it'\sqrt{\frac{12}{r-1}}(x-\frac{1}{2})}\, dx\right]^{r-1} e^{it''\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\, dv$$

$$(5.8) \qquad = \int_0^1 \left[\frac{\sin\left(t'\sqrt{\frac{3}{r-1}}\right)}{t'\sqrt{\frac{3}{r-1}}}\right]^{r-1} e^{it''\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\, dv.$$

Since for $r = \lambda n$ and $n \to \infty$ we have

$$(5.9) \qquad E(v) = \frac{r}{n+1} \to \lambda \quad \text{and} \quad \sigma(v) = \frac{1}{n+1}\sqrt{\frac{r(n - r + 1)}{n + 2}} = O\left(\frac{1}{\sqrt{n}}\right),$$

then we shall write $v = \lambda + O(1/\sqrt{n})$ in the expression (5.6) for t' which is needed for the first part of the integrand in (5.8). For $m = \gamma n$ we obtain the two asymptotic relations

Thus we have for the jth moment

$$(4.3)\qquad E(y^j) = \frac{(-1)^j j!n!}{(r-1)!(n+j)!}[\Delta^{r-1}\{(x-m)^{r-1+j}\}]_{x=0}$$

$$(4.4)\qquad = \frac{(-1)^j j!n!}{(r-1)!(n+j)!}\sum_{\beta=0}^{j}(-1)^\beta m^\beta \binom{r-1+j}{\beta}[\Delta^{r-1}x^{r-1+j-\beta}]_{x=0}.$$

It can be shown that for $j \geqq 0$ and $r \geqq 1$

$$(4.5)\qquad [\Delta^{r-1}x^{r+j}]_{x=0} = \sum_{\alpha=1}^{r-1}\frac{(r-1)!}{(\alpha-1)!}[\Delta^\alpha x^{\alpha+j}]_{x=0}.$$

The results for various values of j in (4.5) are known and are given, for example, in [4], p. 127. Using these we have from (4.4)

$$(4.6)\qquad E(y) = \frac{r(2m-r+1)}{(2n+1)},$$

$$(4.7)\qquad E(y^2) = \frac{r(r+1)}{12(n+1)(n+2)}\left[12m^2 - 12m(r-1) + (r-1)(3r-2)\right],$$

$$(4.8)\qquad \sigma^2(y) = \frac{r(n-r+1)(2m-r+1)^2}{4(n+1)^2(n+2)} + \frac{r(r+1)(r-1)}{12(n+1)(n+2)},$$

$$(4.9)\qquad \begin{aligned} E(y^3) &= \frac{r(r+1)(r+2)}{8(n+1)(n+2)(n+3)} \\ &\quad \cdot\left[8m^3 - 12m^2(r-1) + 2m(r-1)(3r-2) - r(r-1)^2\right], \end{aligned}$$

$$(4.10)\qquad \begin{aligned} E(y^4) &= \frac{r(r+1)(r+2)(r+3)}{2(n+1)(n+2)(n+3)(n+4)} \\ &\quad \cdot\left[2m^4 - 4m^3(r-1) + m^2(r-1)(3r-2) - m(r-1)^2 r \right. \\ &\qquad \left. + \frac{(r-1)(15r^3 - 15r^2 - 10r + 8)}{120}\right]. \end{aligned}$$

Since the computation of cumulants leads to no simplification, they have not been given here; they can be obtained by the usual formulae. It should be mentioned that the above expressions for the moments can also be obtained directly by using the moments of the order statistics.

5. Asymptotic normality of $y = T_{r,m}^{(n)}$. We shall randomize the order of the chance variables $t_1, t_2, \cdots, t_{r-1}$ and thus define new unordered *equi-correlated* and identically distributed chance variables $u_1, u_2, \cdots, u_{r-1}$. Furthermore, if we consider the conditional joint distribution of the u_i given v $(=t_r)$, then we have *independent* chance variables which are uniformly distributed from 0 to v.

Let

(5.1) $$y^* = \frac{y - E(y)}{\sigma(y)} \text{ and } y_v^* = \frac{y - E(y \mid v)}{\sigma(y \mid v)},$$

where $y = T_{r,m}^{(n)} = u_1 + u_2 + \cdots + u_{r-1} + (m - r + 1)v$.
The characteristic function of y^* is given by

(5.2) $$\begin{aligned}\varphi_y^*(t) &= \int_0^1 \int_0^v \cdots \int_0^v \exp\left\{it\left(\frac{y - E(y)}{\sigma(y)}\right)\right\}\left[\prod_{i=1}^{r-1} \frac{du_i}{v}\right] g(v)\,dv \\ &= \int_0^1 \int_0^v \cdots \int_0^v \exp\left\{it\,\frac{\sigma(y \mid v)}{\sigma(y)}\left[\frac{y - E(y \mid v)}{\sigma(y \mid v)}\right]\right. \\ &\qquad \left. + it\,\frac{\sigma(v)}{\sigma(y)}\left[\frac{E(y \mid v) - E(y)}{\sigma(v)}\right]\right\}\left[\prod_{i=1}^{r-1} \frac{du_i}{v}\right] g(v)\,dv,\end{aligned}$$

where

(5.3) $$E(y \mid v) = (r - 1)\frac{v}{2} + (m - r + 1)v = (2m - r + 1)\frac{v}{2},$$

(5.4) $$\sigma(y \mid v) = v\sqrt{\frac{r - 1}{12}}; \qquad \sigma(v) = \frac{1}{(n + 1)}\sqrt{\frac{r(n - r + 1)}{n + 2}},$$

and

(5.5) $$g(v) = \frac{n!}{(r - 1)!(n - r)!} v^{r-1}(1 - v)^{n-r}.$$

Letting

(5.6) $$t' = t\,\frac{\sigma(y \mid v)}{\sigma(y)}, \qquad t'' = t\,\frac{\sigma(v)}{\sigma(y)}\left[m - \left(\frac{r - 1}{2}\right)\right],$$

and $x_i = u_i/v\ (i = 1, 2, \cdots, r - 1)$, we obtain

(5.7) $$\varphi_y^*(t) = \int_0^1 \left[\int_0^1 e^{it'\sqrt{\frac{12}{r-1}}(x-\frac{1}{2})}\,dx\right]^{r-1} e^{it''\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\,dv$$

(5.8) $$= \int_0^1 \left[\frac{\sin\left(t'\sqrt{\frac{3}{r - 1}}\right)}{t'\sqrt{\frac{3}{r - 1}}}\right]^{r-1} e^{it''\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\,dv.$$

Since for $r = \lambda n$ and $n \to \infty$ we have

(5.9) $$E(v) = \frac{r}{n + 1} \to \lambda \quad \text{and} \quad \sigma(v) = \frac{1}{n + 1}\sqrt{\frac{r(n - r + 1)}{n + 2}} = O\left(\frac{1}{\sqrt{n}}\right),$$

then we shall write $v = \lambda + O(1/\sqrt{n})$ in the expression (5.6) for t' which is needed for the first part of the integrand in (5.8). For $m = \gamma n$ we obtain the two asymptotic relations

$$(5.10)\qquad \frac{\sigma(y\mid v)}{\sigma(y)} \cong \frac{v}{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2+\lambda^2}} = \frac{\sqrt{\lambda^2}}{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2+\lambda^2}} + O\left(\frac{1}{\sqrt{n}}\right)$$

and

$$(5.11)\qquad \frac{\sigma(v)}{\sigma(y)}\left[m - \left(\frac{r-1}{2}\right)\right] \cong \frac{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2}}{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2+\lambda^2}},$$

so that if we denote the first term in the right hand members of (5.10) and (5.11) by a and b respectively, then $a^2 + b^2 = 1$. Taking the limit in (5.8) as $n \to \infty$ with $r = \lambda n$, $m = \gamma n$ and using the Lebesgue theorem, we can bring the limit operator under the integral sign. Then, using (5.10), we obtain

$$(5.12)\qquad \varphi_y^*(t) \cong \int_0^1 \lim\left[1 - \frac{a^2t^2}{2(r-1)} + O\left(\frac{1}{n^{3/2}}\right)\right]^{r-1} \lim e^{it\left[\frac{r-E(r)}{\sigma(r)}\right]} g(v)\, dv$$

$$(5.13)\qquad = e^{-\frac{a^2t^2}{2}} \int_0^1 \lim e^{it\left[\frac{r-E(r)}{\sigma(r)}\right]} g(v)\, dv.$$

Using the same Lebesgue theorem the limit operator can be taken outside the integral sign. Then, using a result on the asymptotic normality of quantiles given on page 369 in Cramér [2], we obtain

$$(5.14)\qquad \varphi_y^*(t) \cong e^{-\frac{a^2t^2}{2}} e^{-\frac{b^2t^2}{2}} = e^{-\frac{t^2}{2}}$$

since $a^2 + b^2 = 1$. This proves the asymptotic normality of y for $r = \lambda n$, $m = \gamma n$ (γ and λ fixed with $0 < \lambda \leqq 1$ and $\lambda \leqq \gamma < \infty$) and $n \to \infty$.

It should be noted that the above proof holds no matter how fast m tends to infinity. If $m/n \to \infty$ then $a = 0$ and $b = 1$ and (5.14) still holds.

6. Illustration of rapidity of approach to normality. To illustrate the rapidity of approach to normality of the statistics, we shall use the Edgeworth series expansion

$$(6.1)\qquad F_{r,m}^{(n)}(x) = \{\Phi(x)\} - \left\{\frac{1}{3!}\frac{\mu_3}{\sigma^3}\Phi^{(3)}(x)\right\} + \left\{\frac{1}{4!}\left(\frac{\mu_4}{\sigma^4} - 3\right)\Phi^{(4)}(x) + \frac{10}{6!}\left(\frac{\mu_3}{\sigma^3}\right)^2 \Phi^{(6)}(x)\right\} + \cdots,$$

where $\Phi(x)$ is the standard normal c.d.f., $\Phi^{(r)}(x)$ is its rth derivative and x denotes the standardized variate corresponding to t. We wish to compute one, two, and three terms of (6.1) as indicated by the braces for the two special cases of $T_{r,m}^{(n)}$; viz., (i) $m = r$ and (ii) $m = n$. These have been computed for $n = 10$, $r = 5$ and the results are compared in Table I below with the exact values computed from (2.6).

TABLE I

Comparison of exact probability $P[T_{r,n}^{(2)} \leq t]$ and Edgeworth approximations

Case	t	x	Approximations: 1 term	2 terms	3 terms	Exact Probability
(i)	1.5	0.26656	.6051	6340	6418	6427
$r = 5$	2.0	1.21393	.8932	8851	8840	8840
$m = 5$	2.5	2.22131	9868	.9761	9780	9769
$n = 10$	3.0	3.19868	9993	9975	9969	9974
(ii)	4.0	0.30751	6208	.6312	6201	6250
$r = 5$	5.0	1 15329	.8756	8730	8681	8671
$m = 10$	6.0	1.99902	9772	9723	9711	9729
$n = 10$	7.0	2.84475	.9978	9963	9970	9970

7. Acknowledgment. The authors wish to thank Prof. J. W. Tukey for his helpful comments and suggestions.

REFERENCES

[1] D. E. Barton and F. N. David, "Sums of ordered intervals and distances," *Mathematika*, Vol. 2 (1955), pp 150–159
[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, [illegible]
[3] B. Epstein and M. Sobel, "Some theorems relevant to life testing from an exponential distribution," *Ann Math Stat*, Vol 25 (1954), pp 373 [illegible]
[4] L. M. Milne-Thomson, *The Calculus of Finite Differences*, MacMillan London, [illegible]

DETERMINING SAMPLE SIZE FOR A SPECIFIED WIDTH CONFIDENCE INTERVAL

By Franklin A. Graybill

Oklahoma State University

1. Introduction. If an experimenter decides to use a confidence interval to locate a parameter, he is concerned with at least two things: (1) Does the interval contain the parameter? (2) How wide is the interval? In general the answer to these questions cannot be given with absolute certainty, but must be given with a probability statement. If we let α be the probability that the interval contains the parameter, and let β^2 be the probability that the width is less than d units, then the general procedure is to fix α in advance and compute β^2. The value of β^2 is in general a function of the positive integer n, the sample size by which the confidence interval is computed. (β^2 is also a function of α). In most confidence intervals, β^2 increases as n increases. For any particular situation β^2 may be too low to be useful, hence an experimenter may wish to increase β^2 by taking more observations (increasing n). The problem the experimenter then faces is the determination of n such that (A) the probability will be equal to α that the confidence interval contains the parameter, and (B) the probability will be equal to β^2 that the width of the confidence interval will be less than d units (where α, β^2, and d are specified).

To solve this problem will generally require two things: (1) The form of the frequency function from which the sample of size n is to be selected; (2) Some previous information on the unknown parameters in the frequency function.

This suggests that the sample be taken in two steps; the first sample will be used to determine the number of observations to be taken in the second sample so that (A) and (B) will be satisfied.

For a confidence interval on the mean of a normal population with unknown variance this problem has been solved by Stein [1] for $\beta^2 = 1$.

The purpose of this paper is to determine n, to satisfy (A) and (B) for distributions other than the normal.

2. Theory. Suppose X is the width of a confidence interval on a parameter μ with confidence coefficient α. Suppose further that it is desired that the probability be β^2 that X be less than d. The problem is to determine n, the number of observations, on which to base X. Since n depends on the random variables used in step one, n is a random variable.

We will prove the following (we will use the notation $P(A)$ for the probability that the event A occurs):

Theorem. *Let the chance variable X be the width of a confidence interval on a parameter μ based on a sample of size n. Suppose that X depends on n and on an unknown parameter θ (θ may be the parameter μ). Suppose also that there exists a*

Received April 29, 1957; revised October 8, 1957.

function of X, θ, and n, say $g(X; \theta, n)$, such that if $Y = g(X; \theta, n)$, then the distribution of Y does not depend on any unknown parameters except n. Let $f(n)$ be a function of n such that

(1) $$P[Y < f(n)] = \beta \qquad \text{for any} \qquad 0 < \beta < 1.$$

Let the solution of the equation $g(x; \theta, n)$, $= f(n)$ for x be $x = h(\theta, n)$, and suppose the following are true for $x > 0$:

(2)
(a) *$g(x; \theta, n)$ is monotonic increasing in x for every n and θ.*
(b) *$h(\theta, n)$ is monotonic increasing for every n.*
(c) *$h(\theta, n)$ is monotonic decreasing in n for every θ.*
(d) *z is random variable which is available from step one of the procedure such that $P[t(z) > \theta] = \beta$ for $0 < \beta < 1$, where $t(z)$ is a function of z which does not depend on any unknown parameters or on n.*

Let d and β be specified in advance. Then if n is such that the equation

(3) $$h[t(z), n] \leqq d$$

is satisfied ($t(z)$ is known) then the following inequality is true:

(4) $$P(X \leqq d) \geqq \beta^2.$$

PROOF. Substituting into Eq. (1) we get

(5) $$P[g(X, \theta, n) < f(n)] = \beta.$$

Solving for X and using 2(a) gives us

(6) $$P[X < h(\theta, n)] = \beta.$$

For any $\theta_1 \geqq \theta$ we can use 2(b) and obtain

(7) $$P[X < h(\theta_1, n) \mid \theta_1 \geqq \theta] \geqq P[X < h(\theta, n)] = \beta.$$

By considering the joint distribution of X and $t(z)$ we can write

(8) $$P(X \leqq d) \geqq P[X \leqq d, t(z) > \theta] = P[X \leqq d \mid t(z) > \theta] \cdot P[t(z) > \theta].$$

If n is any integer satisfying

(9) $$h[t(z), n] \leqq d,$$

we can use (7), 2(c), and 2(d) in Eq. (8) and obtain

$$P(X \leqq d) \geqq \beta^2.$$

If the function in 2(b) is monotonic *decreasing*, then the theorem is also true but the inequality in 2(d) must be reversed. The theorem is also true if the function in 2(a) is monotonic *decreasing*. The conditions (2) may appear quite stringent; however, many of the functions in common use in statistics satisfy these conditions.

3. Illustrations

Example 1. Suppose we want an α confidence interval on the variance σ^2 of a normal population to be less than d units in length with probability of β^2.

We will define

$$s_n^2 = \frac{1}{n-1} \sum_{1=1}^{n} (v_i - \bar{v})^2,$$

where v_i is distributed normally with mean μ and variance σ^2. An α confidence interval on σ^2 is given by

$$P\left[\frac{(n-1)\, s_n^2}{\chi_1^2(n)} \leqq \sigma^2 \leqq \frac{(n-1)\, s_n^2}{\chi_2^2(n)}\right] = \alpha,$$

where $\chi_1^2(n)$ and $\chi_2^2(n)$ are such that

$$\int_0^{\chi_2^2(n)} W(\chi^2; n)\, d\chi^2 = \frac{1-\alpha}{2},$$

$$\int_{\chi_1^2(n)}^{\infty} W(\chi^2; n)\, d\chi^2 = \frac{1-\alpha}{2},$$

where $W(\chi^2; n)$ is a Chi-square frequency function with $n - 1$ degrees of freedom. The width of the interval is

$$X = (n-1)\, s_n^2\left[\frac{1}{\chi_2^2(n)} - \frac{1}{\chi_1^2(n)}\right].$$

If we let

$$\frac{1}{\chi_2^2(n)} - \frac{1}{\chi_1^2(n)} = C_n,$$

we have $g(X; \sigma^2, n) = X/\sigma^2 C_n = Y$, and we see that Y is distributed as $W(\chi^2; n)$ and is independent of any unknown parameters except n.

Also $f(n)$ is given by $\int_0^{f(n)} W(\chi^2; n)\, d\chi^2 = \beta$, and $h(\theta, n) = \sigma^2 C_n f(n)$.

Suppose in step one of our procedure we observe $u_1, u_2, \cdots, u_m$ which is a random sample of size m from a normal population with variance σ^2. If we let

$$z = \sum_{i=1}^{m} (u_i - \bar{u})^2,$$

then since z/σ^2 is distributed as $W(\chi^2; m)$, it is clear that $P[(z/\sigma^2) > f_m] = \beta$, where f_m is such that

$$\int_{f_m}^{\infty} W(\chi^2; m)\, d\chi^2 = \beta.$$

Hence $t(z) = z/f_m$, and since all the conditions in (2) are satisfied, the sample size for the desired length of the confidence interval is the smallest integral value of n satisfying

$$\frac{f(n) \cdot C_n \cdot z}{f_m} \leqq d .$$

Example 2. Next, suppose it is desired to determine the sample size such that an α confidence interval on the mean of a normal population will have width less than d with probability β^2. Let $v_1, v_2, \cdots, v_n$ be a random sample of size n(to be determined) from a normal distribution with mean μ and variance $\sigma^2 = \theta^2$. If we let

$$s_n^2 = \frac{1}{n-1} \sum (v_i - \bar{v})^2 ,$$

then an α confidence interval on μ is

$$\bar{v} - \frac{t_0 s_n}{\sqrt{n}} \leqq \mu \leqq \bar{v} + \frac{t_0 s_n}{\sqrt{n}} ,$$

where t_0 is such that

$$\int_{t_0}^{\infty} U(t, n)\, dt = \frac{1-\alpha}{2} ,$$

where $U(t, n)$ is "Student's" distribution with $(n - 1)$ degrees of freedom. The length of the interval is

$$X = 2 \frac{t_0 s_n}{\sqrt{n}} .$$

If we let

$$Y = g(X, \theta, n) = \frac{(n-1)s_n^2}{\sigma^2} = \frac{n(n-1)X^2}{4t_0^2 \sigma^2} ,$$

then Y is distributed as a Chi-square variate with $(n - 1)$ degrees of freedom, and is independent of any unknown parameters except n.

If $W(\chi^2; n)$ is a Chi-square frequency function with $(n - 1)$ degrees of freedom, then $f(n)$ is given by

$$\int_0^{f(n)} W(\chi^2; n)\, d\chi^2 = \beta,$$

and

$$X = h(\theta, n) = 2t_0 \sigma \frac{\sqrt{f(n)}}{\sqrt{n(n-1)}} .$$

Suppose $u_1, u_2, \cdots, u_m$ is a random sample of size m from a normal population with variance σ^2 which is available from step one of our procedure.

If we let

$$z = \sum_{i=1}^{m} (u_i - \bar{u})^2 ,$$

then we have $P[(z/\sigma^2) > f_m] = \beta$, where f_m is such that $\int_{f_m}^{\infty} W(\chi^2; m)\, d\chi^2 = \beta$ Hence, $t(z) = (z/f_m)^{\frac{1}{2}}$, and since all the conditions in (2) are satisfied, the sample size for step two is the smallest integral value of n satisfying

$$\frac{2t_0\sqrt{z}}{\sqrt{f_m}} \cdot \frac{\sqrt{f(n)}}{\sqrt{n(n-1)}} \leqq d.$$

It is interesting to compare the method in this paper with the method presented by Stein [1] for setting a confidence interval on the mean of a normal population with unknown variance.

The procedure presented by Stein is to select a two step sample. Suppose the sample in the first step is $u_1, u_2, \cdots, u_m$ and is taken from a normal population with mean μ and variance σ^2. An α confidence interval on μ is

$$\bar{u} - \frac{t_m s}{\sqrt{m}} \leqq \mu \leqq \bar{u} + \frac{t_m s}{\sqrt{m}},$$

where $s^2 = 1/(m-1) \sum (u_i - \bar{u})^2$ and t_m is the appropriate value from "Students" distribution with $m - 1$ degrees of freedom. The width of the interval is $2t_m s/m^{\frac{1}{2}}$ and if this is less than the desired width d, no second step is required. If $2t_m s/m^{\frac{1}{2}} > d$, then n additional observations $w_1, w_2, \cdots, w_n$ are taken where $n \geqq (4t_m^2 s^2/d^2) - m$, and the α confidence interval is

$$\bar{z} - \frac{t_m s}{\sqrt{m+n}} \leqq \mu \leqq \bar{z} + \frac{t_m s}{\sqrt{m+n}},$$

where

$$\bar{z} = \frac{n\bar{w} + m\bar{u}}{m+n}.$$

The width of the interval is $2t_m s/(m+n)^{\frac{1}{2}}$, and this is less than d.

It is to be noted that observations in the second sample are used *only* to compute the mean, $\bar{z}$.

Let us assume that the observations in the first step are taken from a normal population with mean μ_1 and variance σ_1^2, and in the second step the mean is μ_2 and the variance σ_2^2. Stein's method is valid if $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$. However, if $\mu_1 \neq \mu_2$, but $\sigma_1^2 = \sigma_2^2$, the method can still be used to set a specified confidence interval on μ_2; the only alteration is that the second step requires a sample of size $n + m$ and $\bar{z}$ is the mean of this sample. That is to say, the sample mean from step one is not used in computing the interval. In this case if the inequality

$$\frac{4t_m^2 s^2}{d^2} \leqq n$$

in Stein's procedure is compared with the inequality

$$\frac{2t_0\sqrt{z \cdot f(n)}}{\sqrt{f_m \cdot n(n-1)}} \leqq d$$

for the method presented in this paper, it is evident that Stein's procedure is to be preferred.

Next suppose that $\mu_1 \neq \mu_2$ and $\sigma_1^2 \neq \sigma_2^2$. Then Stein's procedure gives a confidence interval on μ_2 with *known* probability (equal to 1) of a specified *width* but the confidence coefficient is *not known*. The method presented in this paper will give a confidence interval on μ_2 with unknown probability of a specified width, but with known confidence coefficient.

Therefore, there may be cases when an experimenter would prefer the method in this paper over the one given by Stein for the mean of a normal distribution.

REFERENCE

[1] CHARLES STEIN, "A two sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol 16 (1945), pp 243-258.

NOTES

AN EXTENSION OF THE OPTIMUM PROPERTY OF THE SEQUENTIAL PROBABILITY RATIO TEST

By M. A. Girshick[1]

Stanford University

Let $f(x, \theta)$ be a family of densities or discrete probability functions dependi on the parameter θ. Let H_0 be the hypothesis $\theta = \theta_0$ and H_1 the hypothes that $\theta = \theta_1$. A sequential probability ratio test of H_0 versus H_1 is defined two numbers A and B. After drawing the mth observation, sampling is co tinued if

$$B < \prod_{i=1}^{m} \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} < A, \tag{1}$$

where $x_1, \cdots, x_m$ are the first m observations. If the probability ratio is least equal to A, H_1 is accepted, and if it is not greater than B, H_0 is accepte

For any sequential procedure T, let the operating characteristic be

$$L(\theta, T) = \Pr\{\text{Accepting } H_0 \mid \theta, T\}, \tag{2}$$

and let $\mathcal{E}_\theta(n \mid T)$ be the expected number of observations required by T wh sampling from $f(x, \theta)$. The so-called optimum property (see [5], for instanc of a sequential probability ratio test, say T^*, is that if $L(\theta_0, T) \geqq L(\theta_0, T$ and $L(\theta_1, T) \leqq L(\theta_1, T^*)$, then

$$\mathcal{E}_{\theta_0}(n \mid T) \geqq \mathcal{E}_{\theta_0}(n \mid T^*), \qquad \mathcal{E}_{\theta_1}(n \mid T) \geqq \mathcal{E}_{\theta_1}(n \mid T^*).$$

In many cases this optimum property can be extended to all values of t parameter. Suppose $\theta_0 < \theta_1$, and let $\bar{\theta}$ be a number to be defined later such th $\theta_0 < \bar{\theta} < \theta_1$. Under conditions stated below, we give the extended optimu property. If

$$\begin{aligned} L(\theta, T) &\geqq L(\theta, T^*), \qquad \theta < \bar{\theta}, \\ L(\theta, T) &\leqq L(\theta, T^*), \qquad \theta > \bar{\theta}, \end{aligned} \tag{3}$$

for all $\theta \neq \bar{\theta}$, then

$$\mathcal{E}_\theta(n \mid T) \geqq \mathcal{E}_\theta(n \mid T^*) \tag{4}$$

Received May 9, 1956; revised November 15, 1957.

[1] The result reported in this note was mentioned by the late M. A. Girshick to sever of his colleagues, but was unpublished at the time of his death. Since I think the resu is of sufficient interest to be in the literature, I have taken the liberty of writing this no in Girshick's name. T. W. Anderson.

The research was sponsored by the Office of Naval Research.

for all θ. Inequalities (3) indicate the premise that T is everywhere as good as T^* in the sense that the operating characteristic for T is at least as high as for T^* for θ on one side of $\bar{\theta}$ and is as least as low as for T^* on the other side of $\bar{\theta}$ Then T^* is everywhere as good as T in terms of expected number of observations.

To demonstrate the property we assume that for $\theta \neq \bar{\theta}$ there is a unique nonzero root, say $h(\theta)$, of

$$\text{(5)} \qquad \mathcal{E}_\theta \left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^h = 1,$$

and that $h(\theta) > 0$ for $\theta < \bar{\theta}$ and $h(\theta) < 0$ for $\theta > \bar{\theta}$ (See [4] for discussion of the assumption and of the technique used here) This implies that given θ_0 and θ_1 the value of $\bar{\theta}$ for which the assumption holds is unique We make the further assumption that for each θ there is a θ' such that

$$\text{(6)} \qquad \left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^{h(\theta)} f(x, \theta) = f(x, \theta').$$

We now prove (4) for $\theta < \bar{\theta}$ by assuming (3) for θ and θ'. Since

$$h(\theta') = -h(\theta),$$

we have $\theta' > \bar{\theta}$. The sequential probability ratio test T^* defined by (1) can also be defined by

$$\text{(7)} \qquad B^{h(\theta)} < \prod_{i=1}^{m} \left[\frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}\right]^{h(\theta)} < A^{h(\theta)}$$

or by

$$\text{(8)} \qquad B^{h(\theta)} < \prod_{i=1}^{m} \frac{f(x_i, \theta')}{f(x_i, \theta)} < A^{h(\theta)}.$$

Then (4) follows by the usual optimum property because T^* is a sequential probability ratio test for testing hypothesis θ versus the hypothesis θ'. For $\theta > \bar{\theta}$ a similar argument can be used.

The conditions assumed for this extended property are satisfied by many distributions. In particular the existence of such so-called conjugate pairs for distributions of the Koopman-Darmois form has been shown [2]. Savage [3] has shown that the assumptions restrict the families to have a certain exponential form (which includes the Koopman-Darmois form). This note makes explicit Blasbalg's statement [1] that a sequential probability ratio test is optimum at an infinity of parameter points.

REFERENCES

[1] H. Blasbalg, "Sequential analysis," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 1024–1028.
[2] M. A. Girshick, "Contributions to the theory of sequential analysis II, III," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 282–298

[3] L. J. SAVAGE, "When different pairs of hypotheses have the same family of likelihood-ratio test regions," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 1028–1032.
[4] A. WALD, *Sequential Analysis*, John Wiley and Sons, New York, 1947.
[5] A. WALD AND J. WOLFOWITZ, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326–339.

A NOTE ON BALANCED DESIGNS

BY V. R. RAO

University of Bombay

0. Summary. It is proved that a necessary and sufficient condition for a general design to be balanced is that the matrix of the adjusted normal equations for the estimates of treatment effects has $v - 1$ equal latent roots other than zero.

1. Estimates and their properties. We consider a design whose incidence matrix is $N_{v \times b} = [n_{ij}]$ in which the ith treatment is replicated r_i times and the blocks are of sizes $k_1, \cdots, k_b$. With the usual assumptions, the adjusted normal equations for the treatment effects are

$$Q = C\hat{\tau}, \tag{1.1}$$

where

$$Q = T - N \operatorname{diag}\left(\frac{1}{k_1}, \cdots, \frac{1}{k_b}\right) B \tag{1.2}$$

and

$$C = \operatorname{diag}(r_1, \cdots, r_v) - N \operatorname{diag}\left(\frac{1}{k_1}, \cdots, \frac{1}{k_b}\right) N' \tag{1.3}$$

with the condition

$$E_{1v}\hat{\tau} = 0 \tag{1.4}$$

(where E_{pq} denotes a $p \times q$ matrix with all its elements as unity).

It is well known that if rank $C = v - t$, a set of $t - 1$ independent treatment contrasts are not estimable. But if rank $C = v - 1$ every contrast is estimable and in this case the design is said to be connected.

If the design is connected there are $v - 1$ non-zero latent roots, say, $\lambda_1, \lambda_2, \cdots, \lambda_{v-1}$. As the rows of C add to zero, $(v^{-1/2}, \cdots, v^{-1/2})$ is the latent vector corresponding to the root zero.

Let

$$L = \left[\frac{L_1}{v^{-\frac{1}{2}}E_{1v}}\right] = \left[\frac{(l_{ij})}{v^{-\frac{1}{2}}E_{1v}}\right] \tag{1.5}$$

be an orthogonal matrix transforming C into diagonal form.

Received February 28, 1957; revised July 15, 1957.

Since L is orthogonal,

$$I = L'L = L_1'L_1 + \frac{1}{v} E_{vv}. \tag{1.6}$$

Pre-multiplying (1.1) by L, we get

$$LQ = LC\hat{\tau} = \begin{pmatrix} \text{diag } (\lambda_1, \cdots, \lambda_{v-1}) \\ 0 \end{pmatrix} L_1 \hat{\tau}. \tag{1.7}$$

Hence we get

$$L_1 \hat{\tau} = \text{diag} \left(\frac{1}{\lambda_1}, \cdots, \frac{1}{\lambda_{v-1}} \right) L_1 Q.$$

Premultiplying by L_1' and using (1.6) and (1.4), we obtain

$$\hat{\tau} = DQ, \tag{1.8}$$

Where

$$D = [d_{ij}] = L_1' \text{ diag} \left(\frac{1}{\lambda_1}, \cdots \cdots, \frac{1}{\lambda_{v-1}} \right) L_1. \tag{1.9}$$

From the solution (1.8), it follows that

$$\begin{aligned} V(\hat{\tau}) &= D\sigma^2, \\ V(\hat{\tau}_i - \hat{\tau}_j) &= (d_{ii} + d_{jj} - 2\, d_{ij})\sigma^2 \\ &= \sum_{\nu=1}^{v-1} \frac{(l_{\nu i} - l_{\nu j})^2}{\lambda_\nu} \sigma^2. \end{aligned} \tag{1.10}$$

$$\begin{aligned} \text{Average variance} &= \frac{1}{v(v-1)} \sum_{i} \sum_{\substack{j=1 \\ i \neq j}}^{v} V(\hat{\tau}_i - \hat{\tau}_j) \\ &= \frac{2\sigma^2}{v-1} \sum_{\nu=1}^{v-1} \frac{1}{\lambda_\nu} \end{aligned} \tag{1.11}$$

in view of the orthogonality conditions, a result which was obtained by O. Kempthorne in an alternative way [1].

DEFINITION. A design is said to be balanced if every elementary contrast, $\tau_i - \tau_j$ is estimated with the same variance.

2. Theorem. *A necessary and sufficient condition for a design to be balanced is that C has $v - 1$ equal latent roots other than zero.*

To prove that the condition is necessary it is enough to show that $\lambda_1 = \cdots = \lambda_{v-1}$, for the C matrix of a balanced design is of rank $v - 1$.

From (1.10) and (1.11), we get

$$\sum_{\nu=1}^{v-1} \frac{(l_{\nu i} - l_{\nu j})^2}{\lambda_\nu} = \frac{2}{v-1} \sum_{j=1}^{v-1} \frac{1}{\lambda_\nu} = \frac{1}{v-1} \sum_{i=1}^{v-1} \frac{1}{\lambda_i} \sum_{\nu=1}^{v-1} (l_{\nu i} - l_{\nu j})^2.$$

Hence

$$\sum_{\nu=1}^{v-1} (l_{\nu i} - l_{\nu j})^2 \left(\frac{1}{\lambda_\nu} - \frac{1}{v-1} \sum_{k=1}^{v-1} \frac{1}{\lambda_k}\right) = 0. \tag{2.1}$$

Consider

$$V(\hat{\tau}_j - \hat{\tau}_k) = V(\hat{\tau}_i - \hat{\tau}_j) + V(\hat{\tau}_i - \hat{\tau}_k) - 2 \operatorname{Cov} (\hat{\tau}_i - \hat{\tau}_j, \hat{\tau}_i - \hat{\tau}_k).$$

Hence,

$$\operatorname{Cov} (\hat{\tau}_i - \hat{\tau}_j, \hat{\tau}_i - \hat{\tau}_k) = \frac{1}{v-1} \sum_{\nu=1}^{v-1} \frac{1}{\lambda_\nu},$$

i.e.,

$$\sum_{\nu=1}^{v-1} \frac{(l_{\nu i} - l_{\nu j})(l_{\nu i} - l_{\nu k})}{\lambda_\nu} = \frac{1}{v-1} \sum_{\nu=1}^{v-1} \frac{1}{\lambda_\nu} \sum_{\nu'=1}^{v-1} (l_{\nu' i} - l_{\nu' j})(l_{\nu' i} - l_{\nu' k}).$$

Hence,

$$\sum_{\nu=1}^{v-1} (l_{\nu j} - l_{\nu j})(l_{\nu j} - l_{\nu k}) \left(\frac{1}{\lambda_\nu} - \frac{1}{v-1} \sum_{\nu'=1}^{v-1} \frac{1}{\lambda_{\nu'}}\right) = 0 \quad \text{for } i \neq j \neq k, \tag{2.2}$$

From (2.1) and (2.2) taking $i = 1$, we get

$$d^{(j)'} \operatorname{diag} \left(\frac{1}{\lambda_1} - \frac{1}{v-1} \sum \frac{1}{\lambda_\nu}, \cdots, \frac{1}{\lambda_{v-1}} - \frac{1}{v-1} \sum \frac{1}{\lambda_\nu}\right) d^{(k)} = 0 \tag{2.3}$$
$$\text{for } j, k = 2, 3, \cdots v,$$

where $d^{(j)}$ is the column vector

$$\{l_{11} - l_{1j}, l_{21} - l_{2j}, \cdots, l_{v-11} - l_{v-1j}\}.$$

If

$$M = [d^{(2)}, d^{(3)}, \cdots, d^{(v)}],$$

then

$$M'M = \begin{bmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \cdots & & \cdots & \\ 1 & 1 & \cdots & 2 \end{bmatrix},$$

and det. $M'M = v \neq 0$. Hence $M'M$ and hence M are non-singular.

Therefore $d^{(2)}, \cdots, d^{(v)}$ are $v - 1$ linearly independent $(v - 1)$-vectors. Any $(v - 1)$ vector, say ξ, can be uniquely expressed in terms of these vectors, say

$$\xi = C_2 d^{(2)} + C_3 d^{(3)} + \cdots + C_v d^{(v)}.$$

From (2.3) it follows that

$$(2.4)\quad \begin{aligned} &\xi' \operatorname{diag}\left(\frac{1}{\lambda_1} - \frac{1}{v-1}\sum\frac{1}{\lambda_i}, \cdots, \frac{1}{\lambda_{v-1}} - \frac{1}{v-1}\sum\frac{1}{\lambda_i}\right)\xi = 0 \\ &= \sum_{i,j=2}^{v} C_i C_j d^{(i)'} \operatorname{diag}\left(\frac{1}{\lambda_1} - \frac{1}{v-1}\sum\frac{1}{\lambda_i}, \cdots, \frac{1}{\lambda_{v-1}} - \frac{1}{v-1}\sum\frac{1}{\lambda_i}\right)d^{(j)} = 0. \end{aligned}$$

By taking successively ξ to be the $(v-1)$ vectors $(1, 0, \cdots, 0), \cdots$ and $(0, 0, \cdots, 1)$, we get

$$\frac{1}{\lambda_1} = \frac{1}{\lambda_2} = \cdots = \frac{1}{\lambda_{v-1}}.$$

Hence

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{v-1}$$

The condition is sufficient, for, if rank $C = v - 1$ and $\lambda_1 = \cdots = \lambda_{v-1} = \lambda$ (say), it follows immediately that every elementary contrast is estimable, and the solutions become

$$\hat{\tau} = \frac{1}{\lambda} L_1' L_1 Q = \frac{1}{\lambda}\left(I - \frac{1}{v} E_{vv}\right) Q = \frac{Q}{\lambda},$$

which shows that $V(\hat{\tau}_i - \hat{\tau}_j) = (2/\lambda)\sigma^2$, which is independent of both i and j and hence, the design is balanced. Q.E.D.

COROLLARIES. (i) *If the design is balanced, then*

$$(2.5)\qquad C = \lambda I - \frac{\lambda}{v} E_{vv}$$

and the solutions are

$$(2.6)\qquad \hat{\tau}_j = \frac{Q_j}{\lambda}.$$

(ii) *In a balanced design with equal block sizes, k, the replicate numbers must be equal.*

PROOF. $C = \operatorname{diag}(r_1, \cdots, r_v) - 1/kNN'$ if block size is constant. Hence by Eq. (2.5), if the design is also balanced, we have

$$r_i - \frac{r_i}{k} = \lambda - \frac{\lambda}{v}.$$

Hence, r_i is the same constant for all i. Q.E.D.

(iii) *If all the treatments are replicated the same number of times and the blocks are of the same size then the only balanced design is* BIBD, *if such a design exists.*

PROOF. If r is the number of replications and k is the block size, then

$$(2.7)\quad \begin{aligned} C &= rI - \frac{1}{k} NN' \\ &= \lambda I - \frac{\lambda}{v} E_{vv} \text{ by Corollary (i).} \end{aligned}$$

Hence comparing off-diagonal elements, we get

$$\lambda_{ij} = \frac{k\lambda}{v},$$

where λ_{ij} is the number of times the pair of treatments i, j occur together in the blocks. Since λ_{ij}'s are all equal the design is Balanced Incomplete Block Design (BIBD) [2]. This result was proved in an alternative form by W. A. Thompson [3].

3. Concluding remarks. But these do not exclude the possibilities of the existence of balanced designs with different block sizes and the same number of replications. As an example consider the design whose incidence matrix is

$$N = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix},$$

$$r = (6, 6, 6, 6); \qquad k = (3, 3, 3, 3, 2, 2, 2, 2, 2, 2).$$

Here it can be verified that every elementary contrast is estimated with a variance equal to $3\sigma^2/7$, but the design is not a Balanced Incomplete Block Design.

It can also be seen that the example given above is obtained by adjoining two BIBD's with the same number of treatments. Such designs can be constructed from two BIBD's with the same number of treatments. Investigations on these lines are being carried out.

Acknowledgement. The author wishes to express his indebtedness to Professor M. C. Chakrabarti for suggesting this problem and for his help and guidance in preparing this note.

REFERENCES

[1] O. Kempthorne, "The Efficiency Factor of an Incomplete Block Design," *Ann. Math. Stat.* Vol. 27 (1956), pp. 846–849.

[2] F. Yates, "Incomplete Randomized Blocks," *Annals of Eugenics*, Vol. VII (1937), pp. 121–140.

[3] W. A. Thompson, "A Note on Balanced Incomplete Block Designs," *Ann. Math. Stat.* Vol. 27 (1956), pp. 842–846.

THE SPACING OF OBSERVATIONS IN POLYNOMIAL REGRESSION

By P. G. Guest

University of Sydney, Australia

1. Introduction and summary. De la Garza ([1], [2]) has considered the estimation of a polynomial of degree p from n observations in a given range of the

Received April 8, 1957.

independent variable x. This range may conveniently be taken to be from $+1$ to -1. He showed that for any arbitrary distribution of the points of observation there was a distribution of the n observations at only $p + 1$ points for which the variances (determined by the matrix $X^T W X$) were the same. He then considered how these $p + 1$ points should be distributed so that the maximum variance of the fitted value in the range of interpolation should be as small as possible. In the present note general formulae will be obtained for the distribution of the points of observation and for the variances of the fitted values in the minimax variance case, and the variances will be compared with those for the uniform spacing case.[1]

2. Spacing for minimax variance. The fitted value is given by

$$u_p(x) = \sum_{j=0}^{p} L_j(x)\bar{y}_j, \tag{1}$$

where $L_j(x)$ is the Lagrangian coefficient corresponding to the point of observation x_j and $\bar{y}_j$ is the mean of the observed values at this point. The variance of the fitted value is $\operatorname{var} u_p(x) = \sum_{j=0}^{p} L_j^2(x) \operatorname{var} \bar{y}_j$.

At a point of observation

$$L_j(x_k) = \delta_{jk}$$

and

$$\operatorname{var} u_p(x_j) = \operatorname{var} \bar{y}_j$$

The largest value of this variance will be as small as possible when the n observations are equally divided among the $p + 1$ points. When this is done

$$\operatorname{var} u_p(x_j) = (p + 1)\sigma^2/n \tag{2}$$

and

$$\operatorname{var} u_p(x) = \sum_{j=0}^{p} L_j^2(x)(p + 1)\sigma^2/n. \tag{2.1}$$

Since this is a polynomial of degree $2p$, the minimax variance conditions are obtained when the maxima of $\operatorname{var} u_p(x)$ are at the $p - 1$ internal points x_j, and the end points x_0 and x_p are $+1$ and -1; for then $\operatorname{var} u_p(x)$ never exceeds

$$(p + 1)\sigma^2/n$$

in the range $+1$ to -1. The minimax variance conditions are thus

$$L_j'(x_j) = 0, \quad j = 1 \text{ to } p - 1. \tag{3}$$

---

[1] K. Smith, in an earlier discussion, has given details of curves up to the sixth degree (Biometrika 12 (1918), pp. 1-85).

Now, if

$$F(x) = \prod_{j=0}^{p} (x - x_j), \tag{4}$$

then

$$L_j(x) = \frac{F(x)}{(x - x_j)F'(x_j)},$$

and so

$$F''(x) = \{(x - x_j)L_j''(x) + L_j'(x)\}F'(x_j),$$

and (3) is equivalent to

$$F''(x_j) = 0, \qquad j = 1 \text{ to } p - 1. \tag{5}$$

The function $F(x)$ will be of the form $\alpha(x^2 - 1)\phi_{p-1}(x)$, where the polynomial $\phi_{p-1}(x)$ of degree $p - 1$ is determined by the $p - 1$ equations (5). The polynomial which satisfies these equations is readily shown to be the derivative $P_p'(x)$ of the Legendre polynomial. For if

$$F(x) = \alpha(x^2 - 1)P_p'(x), \tag{6}$$

then

$$F'(x) = \alpha \frac{d}{dx}\{(x^2 - 1)P_p'(x)\} = \alpha p(p + 1)P_p(x)$$

and

$$F''(x) = \alpha p(p + 1)P_p'(x),$$

and so $F''(x)$ vanishes at the internal points $F(x) = 0$.

The points of observation for minimax variance are then to be located at $+1$, -1, and the roots of $P_p'(x) = 0$.

Since the internal points of observation are points of maximum variance, the variance will be given by an equation of the form

$$\text{var } u_p(x) = \{1 + \beta(x^2 - 1)P_p'^2(x)\}(p + 1)\sigma^2/n. \tag{7}$$

The minima of the variance curve then occur at points for which

$$xP_p'(x) + (x^2 - 1)P_p''(x) = 0,$$

and this equation is equivalent to

$$xP_p'(x) = p(p + 1)P_p(x). \tag{8}$$

From (2.1),

$$\text{var } u_p(x) = \sum_{j=0}^{p}\left\{\frac{\alpha(x^2 - 1)P_p'(x)}{(x - x_j)\alpha p(p + 1)P_p(x_j)}\right\}^2 (p + 1)\sigma^2/n, \tag{9}$$

and so, on comparing the coefficients of $x^2 P_p'^2(x)$ in (7) and (9),

$$\beta = \sum_{j=0}^{p} \{p(p+1)P_p(x_j)\}^{-2}.$$

The Lobatto quadrature formula [3] with $f(x) \equiv 1$ gives

$$\int_{-1}^{1} dx = \sum_{j=0}^{p} 2\{p(p+1)P_p^2(x_j)\}^{-1} = 2.$$

Thus the explicit formula for the variance of the fitted value is

$$\text{(10)} \qquad \operatorname{var} u_p(x) = \left\{ 1 + \frac{x^2 - 1}{p(p+1)} P_p'^2(x) \right\} (p+1)\sigma^2/n$$

In the region of extrapolation, when $|x|$ is large

$$P_p'(x) = p\{(2p)!/2^p p!^2\}x^{p-1},$$

and so

$$\text{(11)} \qquad \operatorname{var} u_p(x) = p\{(2p)!/2^p p!^2\}^2 x^{2p} \sigma^2/n$$

3. Uniform spacing. When the observations are spaced at equal intervals the variance of the fitted value is

$$\operatorname{var} u_p(x) = \sum_{j=0}^{p} \left\{ T_j^2(x) \Big/ \sum_{i=1}^{n} T_j^2(x_i) \right\} \sigma^2,$$

where the $T_j(x)$ are the polynomials orthogonal over the n points of observation x_i. When n is large these polynomials will approximate to multiples of the Legendre polynomials $P_j(x)$ which are orthogonal over the continuous range $+1$ to -1. Thus

$$T_j(x) \sim k_j P_j(x)$$

and

$$\sum_i T_j^2(x_i)\Delta x_i \sim k_j^2 \int_{-1}^{1} P_j^2(x)\, dx = 2k_j^2/(2j+1)$$

The interval Δx_i between neighboring observations is $2/n$, and so

$$\sum T_j^2(x_i) \sim nk_j^2/(2j+1)$$

and

$$\text{(12)} \qquad \operatorname{var} u_p(x) \sim \sum_{j=0}^{p} (2j+1)P_j^2(x)\sigma^2/n.$$

The maxima and minima of variance are at points given by

$$\sum_{j=0}^{p} (2j+1)P_j(x)P_j'(x) = 0,$$

which from the recurrence relations for Legendre polynomials is

$$\sum_{j=0}^{p} P_j'(x)\{P_{j+1}'(x) - P_{j-1}'(x)\} = 0,$$

or

$$P_p'(x)P_{p+1}'(x) = 0.$$

The points of maximum variance are then the roots of $P_p'(x) = 0$ and the points of minimum variance the roots of $P_{p+1}'(x) = 0$. It is interesting to observe that the points of observation in the minimax variance method are points of maximum variance in the uniform spacing method. These points are also the points used in the Lobatto quadrature formula.

The Christoffel-Darboux identity [3] for the sum in equation (12) leads to the alternative form

$$\text{(12.1)} \qquad \operatorname{var} u_p(x) \sim \{P_p(x)P_{p+1}'(x) - P_p'(x)P_{p+1}(x)\}(p+1)\sigma^2/n.$$

By use of the recurrence relations for the Legendre polynomials this can be put in the form

$$\text{(12.2)} \qquad \operatorname{var} u_p(x) \sim \left\{(p+1)P_p^2(x) - \frac{x^2-1}{p+1}P_p'^2(x)\right\}(p+1)\sigma^2/n.$$

At the end-points $+1$ and -1, $P_p^2(x)$ is unity and

$$\operatorname{var} u_p(\pm 1) \sim (p+1)^2\sigma^2/n.$$

At the centre of the range the variance can be obtained by substituting the values of $P_p(0)$ and $P_p'(0)$ in (12.2). It is found that

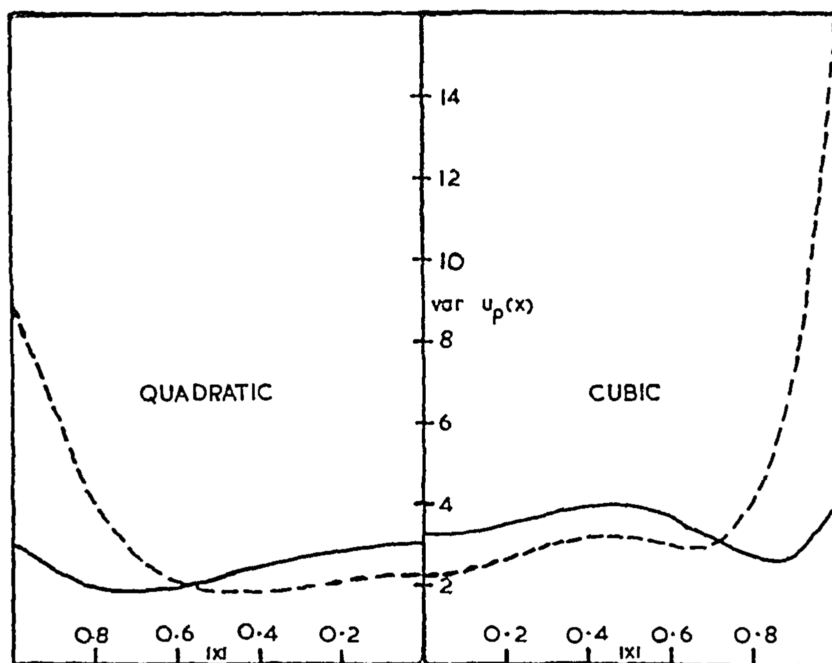


Fig. 1. The solid curve shows the variance of the fitted value for the minimax variance method and the dotted curve the variance for the uniform spacing method. The unit for the variance scale is σ^2/n.

(13) $$\operatorname{var} u_p(0) \sim \left\{\frac{(2q+1)(2q-1)\cdots 1}{2^q q!}\right\}^2 \sigma^2/n,$$

where q is $\frac{1}{2}p$ when p is even and $\frac{1}{2}(p-1)$ when p is odd. In the region of extrapolation, when $|x|$ is large (12 2) gives

$$\operatorname{var} u_p(x) = (2p+1)\{(2p)!/2^p p!^2\}^2 x^{2p}\sigma^2/n$$

The deviations from these formulae when n is not large have been discussed and tabulated [4].

4. Comparison of the two methods. In the central part of the range the uniform spacing method gives a smaller variance than the minimax variance method. An asymptotic expansion of (13) using Stirling's factorial approximation shows that the ratio of the variances is roughly $2/\pi$. This ratio increases steadily with $|x|$, and at the ends of the range the variance for the uniform spacing method exceeds that for the minimax variance method by a factor $p + 1$, while in the region of extrapolation this factor approaches $2 + p^{-1}$. The crossover points for the two variance curves occur at ± 0.58 for the quadratic and ± 0.72 for the cubic Thus over most of the region of interpolation the advantage lies with the uniform spacing method, but at the extremes of the region of interpolation and in the region of extrapolation the advantage lies decidedly with the minimax variance method

Fig. 1 shows the shape of the two variance curves in the region of interpolation for the second and third degree polynomials Since the curves are symmetrical about the origin of x, only half of each curve is drawn.

REFERENCES

[1] A. DE LA GARZA, "Spacing of information in polynomial regression," *Ann Math Stat.*, Vol 25 (1954), pp 123-130
[2] A. DE LA GARZA, D S. HAWKHURST AND L T NEWMAN, "Some minimum cost experimental procedures in quadratic regression," *J Amer Stat Assn*, Vol 50 (1955), pp. 178-184
[3] F. B. HILDEBRAND, *Introduction to Numerical Analysis*, McGraw Hill Book Co, New York, 1956, pp. 319-345
[4] P. G. GUEST, "Estimation of the error at a point on a least squares curve," *Australian J. Sci. Research Ser A*, Vol. 3 (1950) pp 173-182

CONDITIONS THAT A STOCHASTIC PROCESS BE ERGODIC[1]

BY EMANUEL PARZEN

Stanford University

In his work on statistical inference on stochastic processes, Grenander has pointed out ([2], p. 257) that "the concept of metric transitivity seems to be

Received June 3, 1957; revised August 2, 1957

[1] This work was supported in part by the Office of Naval Research Reproduction in whole or in part is permitted for any purpose of the United States Government.

important in the problem of estimation of a stationary stochastic process." In this note, we give necessary and sufficient conditions in terms of characteristic functions that a strictly stationary stochastic process $X(t)$ be metrically transitive or ergodic (see Doob [1], pp. 452–457 for definition of the terminology). More importantly, we state a mean ergodic theorem (or weak law of large numbers) for stochastic processes which are strictly stationary of order K, by which is meant that for every choice of K points $t_1, \cdots, t_K$, the random variables $X(t_1 + h), \cdots, X(t_K + h)$ have a joint probability distribution which does not depend on h.

THEOREM 1. Let the random variables $X(t)$ be defined for t in

$$T = \{0, \pm 1, +2, \cdots\}.$$

Let K be a positive integer. Let $t_1, \cdots, t_K$ be points in T. Assume that there is a characteristic function $\varphi(u_1, \cdots, u_K)$ such that, for all $u_1, \cdots, u_K$,

$$E[\exp i\{u_1 X(t_1 + h) + \cdots + u_K X(t_K + h)\}] = \varphi(u_1, \cdots, u_K) \quad \text{for all } h \text{ in } T. \tag{1.1}$$

Assume that, for each τ in T, there is a characteristic function $\varphi(u_1, \cdots, u_K; \tau)$ such that

$$E[\exp i\{u_1(X(t_1 + h) - X(t_1 + h + \tau)) + \cdots + u_K(X(t_K + h) - X(t_K + h + \tau))\}] = \varphi(u_1, \cdots, u_K; \tau) \quad \text{for all } h \text{ in } T. \tag{1.2}$$

Let $r \geqq 1$. Then for every Borel function $g(x_1, \cdots, x_K)$ such that

$$E\,|\,g(X(t_1), \cdots, X(t_K))|^r < \infty,$$

the sample means

$$M_n(g) = \frac{1}{n+1} \sum_{h=0}^{n} g(X(t_1 + h), \cdots, X(t_K + h)) \tag{1.3}$$

converge as a limit in r-mean. A necessary and sufficient condition that the limit of the $M_n(g)$ be the ensemble mean $E(g) = Eg(X(t_1), \cdots, X(t_K))$ is that, for all real $u_1, \cdots, u_K$,

$$\lim_{n \to \infty} \frac{1}{n+1} \sum_{\tau=0}^{n} \varphi(u_1, \cdots, u_K; \tau) = |\,\varphi(u_1, \cdots, u_K)\,|^2. \tag{1.4}$$

The meaning of these conditions is as follows: (1.1) states, in terms of characteristic functions, that the stochastic process is strictly stationary of order K; (1.2) states that the process of increments $Y(t) = X(t) - X(t + \tau)$ is strictly stationary of order K; (1.4) represents a very weak form of asymptotic independence.

From Theorem 1, together with the Birkhoff-Khintchine ergodic theorem (see Doob [1], pp. 464–473) we immediately obtain the following theorem.

THEOREM 2: A strictly stationary stochastic process $X(t)$ is metrically transi-

tive if, and only if, for every positive integer K, for any choice of K points $t_1, \cdots, t_K$, and for any real numbers $u_1, \cdots, u_K$, (1.4) holds.

The conditions of Theorem 2 constitute a formulation in terms of characteristic functions of known conditions for metric transitivity (see Loève [4], p. 435).

As an indication of the power of these theorems, let us mention that with their aid one can readily establish the following statement made without proof in the book of Grenander and Rosenblatt ([3], p. 44). *If $X(t)$ is a normal process, a necessary and sufficient condition for it to be ergodic (metrically transitive) is that its spectrum be continuous. If $X(t)$ is a linear process, then it is ergodic.*

Theorem 1, and consequently Theorem 2, may be extended to the case of continuous parameter stochastic processes. They provide a new proof of the theorem of Maruyama (see [2], p. 257) that a continuous stationary normal process is metrically transitive if, and only if, its spectrum is continuous.

Theorem 1 is very closely related to the weak law of large numbers for wide-sense stationary processes (see Doob [1], p 489), from which it differs in that it does not require existence of second moments for $X(t)$.

The proof of Theorem 1 is fairly immediate From (1.1), (1.2), and (1.4), it follows (either by the weak law of large numbers for wide-sense stationary processes, or directly by a simple argument [6]) that the theorem holds for trigonometric polynomials $g(x_1, \cdots, x_K) = \exp i(u_1x_1 + \cdots + u_Kx_K)$ To extend the theorem to Borel functions $g(x_1, \cdots, x_K)$ such that $E|g|^r < \infty$, one uses the fact that to any $\epsilon > 0$ one may find a trigonometric polynomial $g_\epsilon(x_1, \cdots, x_K)$ such that

$$E|g(X(t_1), \cdots, X(t_K)) - g_\epsilon(X(t_1), \cdots, X(t_K))|^r < \epsilon.$$

In [5] one may find related theorems, including a discussion of convergence with probability one of certain sample means $M_n(g)$ of stochastic processes which are strictly stationary of order K.

REFERENCES

[1] J. L. Doob, *Stochastic Processes*, New York, 1953
[2] U. Grenander, "Stochastic processes and statistical inference," *Arkiv for Matematik*, Vol. 1 (1950), pp. 195–277.
[3] U. Grenander and M Rosenblatt, *Statistical Analysis of Stationary Time Series*, New York, 1957.
[4] M. Loève, *Probability Theory*, New York, 1955
[5] E. Parzen, "Conditions that a stochastic process be ergodic," Stanford University, *Department of Statistics*, Technical Report No. 11, March, 1957.
[6] E. Parzen, "Laws of large numbers for dependent random variables," to be published as a section in a forthcoming text entitled *An Introduction to Modern Probability Theory*.

POWER FUNCTIONS OF THE GAMMA DISTRIBUTION

By Gerald D. Berndt

Advisory Committee on Weather Control

Power functions are given for testing hypotheses on an increase in the mean μ of a gamma variable.

Let x be a random variable from a gamma population and let the frequency distribution of x be given by

$$f_0 = f(x; \beta, \gamma) = (\beta^\gamma \Gamma(\gamma))^{-1} x^{\gamma-1} \exp(-x/\beta), \qquad x > 0,$$
$$= 0, \qquad x \leqq 0,$$

where $\beta > 0$ and $\gamma > 0$. If x then undergoes a scale change of the form $x \to \delta x$ with $\delta > 1$, it is easily verified that the frequency distribution of δx is given by

$$f_1 = f(\delta x; \delta\beta, \gamma) = ((\delta\beta)^\gamma \Gamma(\gamma))^{-1} x^{\gamma-1} \exp(-x/\delta\beta), \qquad x > 0,$$
$$= 0, \qquad x \leqq 0.$$

Now in testing the null hypothesis $H_0: \mu = \beta\gamma$ against the alternative hypothesis $H_1: \mu = \delta\beta\gamma$, $\delta > 1$ and specified, the probability of detecting the hypothesized change in the mean, or the power of the test, is given by

$$\pi_\delta = \int_{x(\alpha)}^{\infty} f_1 \, dx,$$

where $x(\alpha)$ is such that

$$\alpha = \int_{x(\alpha)}^{\infty} f_0 \, dx$$

and α is the significance level of the test.

Curves of power functions of testing H_0 against H_1 are given for

$$\gamma = \tfrac{1}{2}, 1(1)5, 7, 10(5)50,$$
$$1.0 \leqq \delta \leqq 4.0,$$

and

$$\alpha = 0.01, 0.05, \text{ and } 0.10.$$

For sufficiently large γ, the distribution of x converges to the normal distribution, and for many purposes the power of the test may then be evaluated by simply using the tables of the normal distribution function with standardized variates

$$t_\alpha = (x(\alpha) - \beta\gamma)/\beta\sqrt{\gamma},$$

Received June 10, 1957.

which is exceeded with probability α under H_0, and

$$t_\pi = (t_\alpha\beta\sqrt{\gamma} + \beta\gamma(1 - \delta))/\delta\beta\sqrt{\gamma},$$

which is exceeded with probability π under H_1. The upper bound of the error in using the normal approximation to the gamma distribution for $\gamma \geqq 50$ is calculated, by trial, to be

$$\sup_x |G(x) - N(x)| < 0.019,$$

where $G(x)$ is the distribution function of x as a gamma variable and $N(x)$ is the distribution function of x as a normal variable.

Consider the following example of the use of the accompanying power curves. In illustrating the use of the power curves we first take note of a well known property of the gamma distribution. That is, if $x_i(i = 1, \cdots, n)$ are independent random variables from gamma populations with parameters β and γ_i, then the sample mean is also a gamma variable with parameters β/n and

$$\gamma = \sum_{i=1}^{n} \gamma_i$$

See, for example, [1].

Suppose, for illustration, that a sample of size $n = 10$ is drawn, and that the $x_i(i = 1, \cdots, 10)$ are known to be independently and identically distributed gamma variables with $\gamma_i = 2.0$ and the same β for each i. It is desired to test $H_0: \mu = \mu_0$ against $H_1: \mu = 1.5\mu_0 = \mu_1$ with probability $\alpha = 0.05$ of accepting H_1 when in fact H_0 is true. What is the probability of detecting μ_1? Here $\delta = 1.5$ and $\gamma = 20.0$. In Fig. 2 we find $\delta = 1.5$ on the abscissa and move vertically to the point of intersection with the curve $\gamma = 20.0$. The power, $\pi_{1.5} = 0.598$, is then the ordinate value at this point of intersection.

How large a sample should be drawn in order that there is at least a probability of $\pi_{1.5} = 0.75$ of detecting the specified increase $\delta = 1.5$ in μ_0? Interpolating for the value of γ at $\delta = 1.5$ and $\pi_{1.5} = 0.75$, we find $\gamma = 32$. Hence the sample size should be at least $n = 32/\gamma_i = 16$ in this case.

The calculations on which the power curves are based were made using 3-point Lagrangian interpolation in Pearson's tables of the incomplete gamma function [2]. All calculations have been verified by actual integration of the gamma functions using high-speed computing machinery. This verification was carried out under the supervision of Dr. Max A. Woodbury at New York University.

I wish to acknowledge the work of Dr. Woodbury and his staff in making these calculations, and also to thank Elaine Berndt who performed all of the interpolations we have required and who drafted the accompanying figures.

REFERENCES

[1] C. R. Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952.
[2] *Tables of the Incomplete Gamma Function*, Karl Pearson, editor, Cambridge University Press, re-issued in 1951.

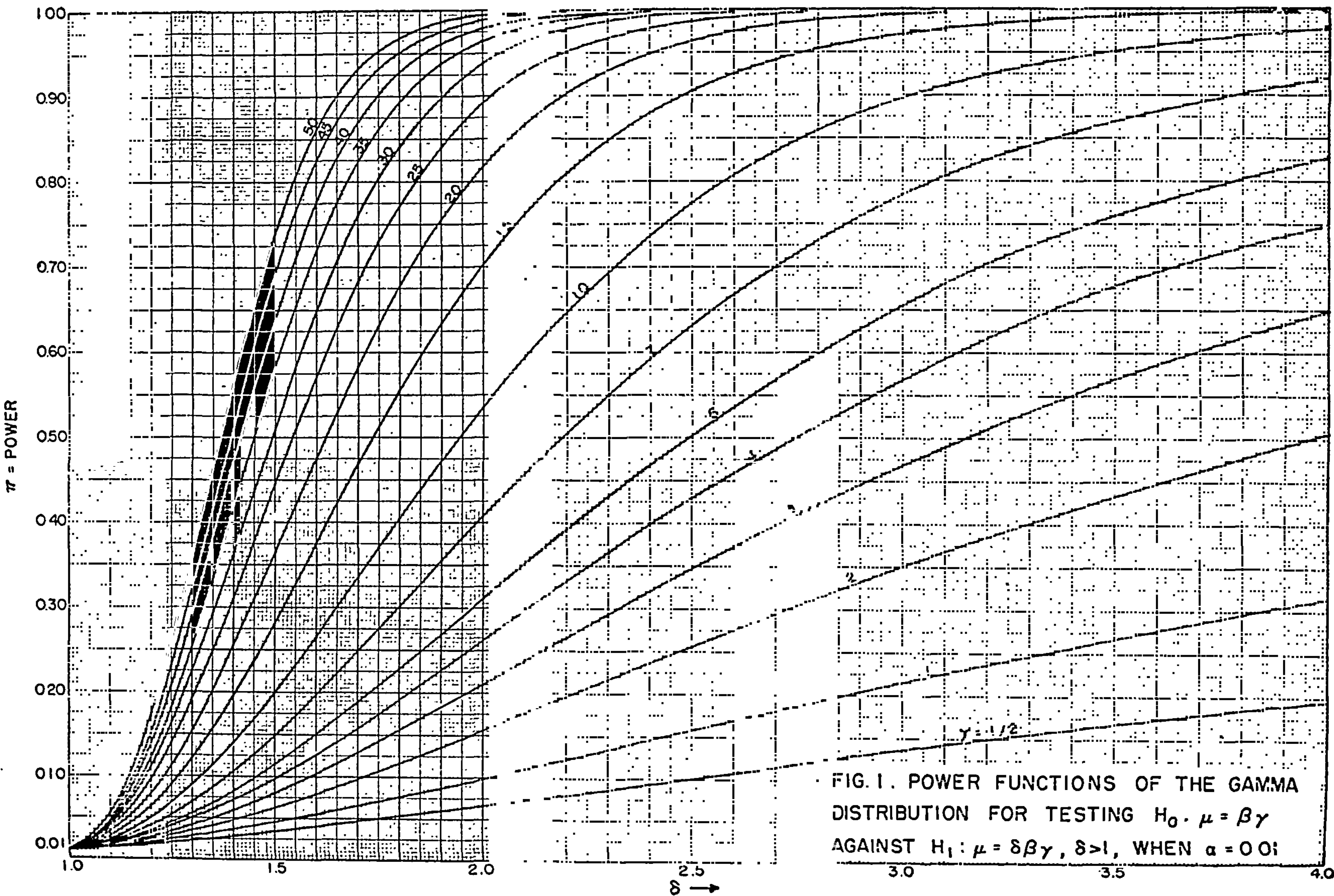
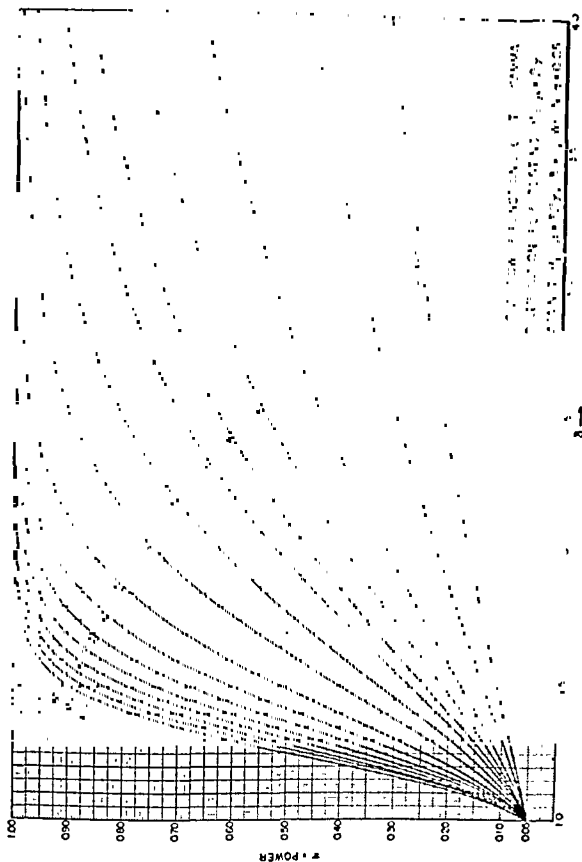


FIG. 1. POWER FUNCTIONS OF THE GAMMA DISTRIBUTION FOR TESTING $H_0: \mu = \beta\gamma$ AGAINST $H_1: \mu = \delta\beta\gamma$, $\delta > 1$, WHEN $\alpha = 0.01$


π = POWER
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00

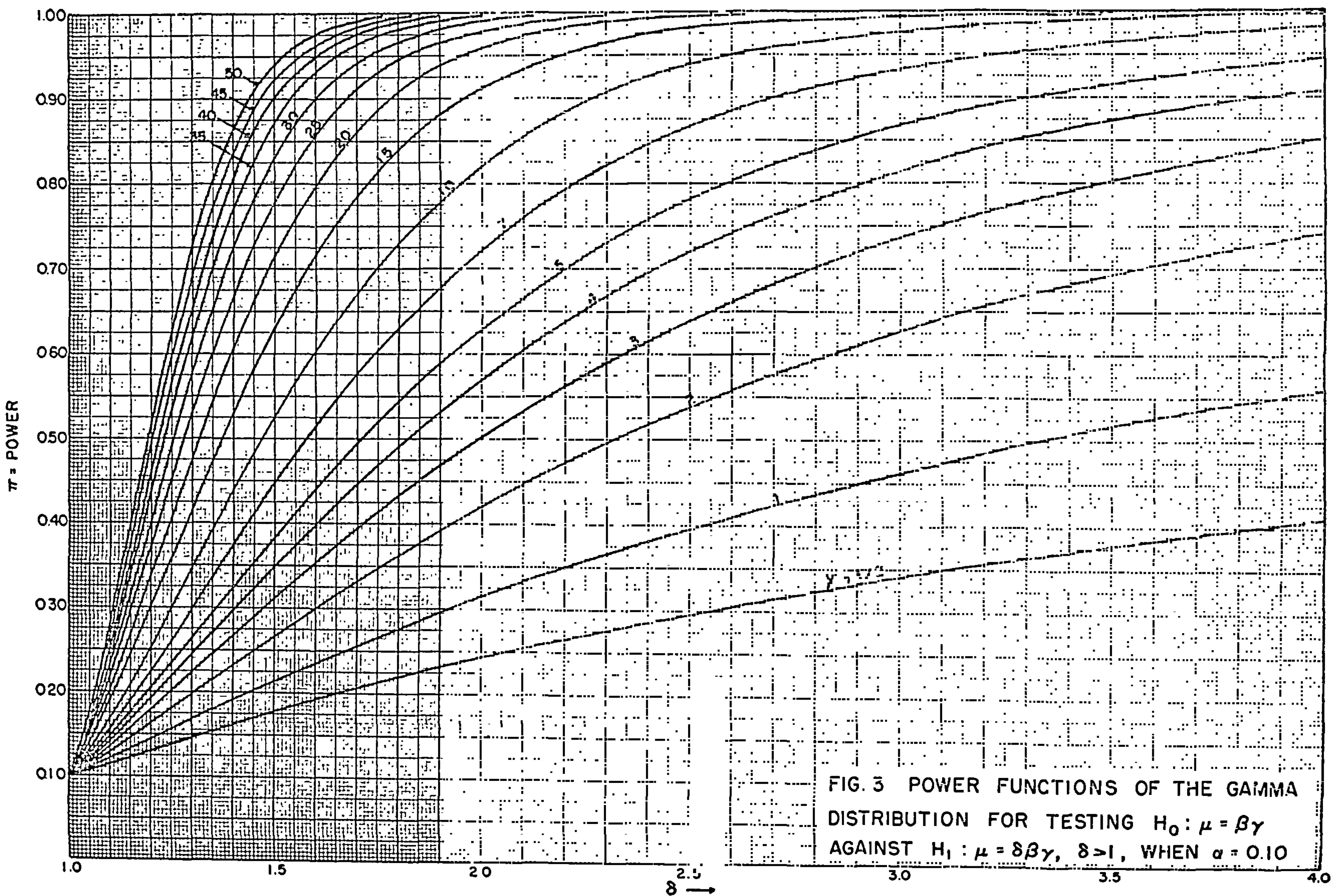


FIG. 3 POWER FUNCTIONS OF THE GAMMA DISTRIBUTION FOR TESTING $H_0: \mu = \beta\gamma$ AGAINST $H_1: \mu = \delta\beta\gamma$, $\delta > 1$, WHEN $\alpha = 0.10$

THE SMALL SAMPLE DISTRIBUTION OF $n\omega_n^2$

By A. W. Marshall

The RAND Corporation

The asymptotic distribution of the statistic

$$(1) \qquad n\omega_n^2 = n \int_{-\infty}^{\infty} [S_n(x) - F(x)]^2 \, dF(x),$$

where $S_n(x)$ is the sample cumulative distribution function (CDF), and $F(x)$ the true CDF, is known and tabled [1] Below are tabled some values of the CDF's of $n\omega_n^2$ for $n = 1, 2$, and 3 Convergence to the asymptotic distribution appears to be extremely rapid.

1. General considerations. It is well-known that: (A) the distribution of $n\omega_n^2$ is distribution free so that it is sufficient to treat the case where $F(x)$ is uniform on the interval [0, 1], and (B) an equivalent form, especially suitable for computation from the ordered sample $x_1 \leqq x_2 \leqq \cdots \leqq x_n$, is

$$(2) \qquad n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i - 1}{2n} - F(x_i)\right]^2$$

or for the case where $F(x)$ is uniform [0, 1]

$$(3) \qquad n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i - 1}{2n} - x_i\right]^2.$$

As was suggested to me several years ago by Oliver Gross (3) clearly shows that the CDF of the $n\omega_n^2$ statistic can be evaluated rather easily for small n. The case $n = 1$ is trivial. For $n = 2$ one must evaluate the area in the intersection of a circle with its center at $x_1 = \frac{1}{4}$, $x_2 = \frac{3}{4}$ and a triangle with vertices at (0, 0), (0, 1), and (1, 1) For $n = 3$ one must evaluate the volume in the intersection of a sphere with center at the point $(\frac{1}{6}, \frac{1}{2}, \frac{5}{6})$ and the tetrahedron with vertices at (0, 0, 0), (0, 0, 1), (0, 1, 1), (1, 1, 1) From (3) one also derives the result that $n\omega_n^2$ has a minimum value of $1/12n$ and a maximum value of $n/3$.

2. Case A: $n = 1$. Since

$$\omega_1^2 = \tfrac{1}{12} + [\tfrac{1}{2} - x_1]^2$$

the CDF of ω_1^2 is

$$F_1(z) = \Pr\{\omega_1^2 \leqq z\} = \begin{cases} 0, & z < \frac{1}{12}, \\ (4z - \frac{1}{3})^{\frac{1}{2}}, & \frac{1}{12} \leqq z \leqq \frac{1}{3}, \\ 1, & z > \frac{1}{3}. \end{cases}$$

3. Case B: $n = 2$.

$$2\omega_2^2 = \tfrac{1}{24} + [\tfrac{1}{4} - x_1]^2 + [\tfrac{3}{4} - x_2]^2.$$

Received June 25, 1957.

By evaluating the area common to a circle of radius $(z - 1/24)^{1/2}$ with center at $(\frac{1}{4}, \frac{3}{4})$ and the triangle with vertices at $(0, 0)$, $(0, 1)$, and $(1, 1)$, multiplying by two, the CDF $F_2(z)$ of the associated value of $2\,\omega_2^2$ is obtained. The result is:

$$
\begin{aligned}
&0, && z < \frac{1}{24},\\
&2\pi\left(z - \frac{1}{24}\right), && \frac{1}{24} \leqq z \leqq \frac{5}{48},\\
&\left(z - \frac{1}{24}\right)\left[2\pi - 4\ \mathrm{Cos}^{-1}\ \tfrac{1}{4}(z - 1/24)^{-1/2} + \frac{(z - 5/48)^{1/2}}{z - 1/24}\right], && \frac{5}{48} < z \leqq \frac{1}{6},\\
&\left(z - \frac{1}{24}\right)\left[\frac{3\pi}{2} - 2\ \mathrm{Cos}^{-1}\ \frac{\tfrac{1}{4}(1/8)^{1/2} - (z - 1/6)^{1/2}(z - 5/40)^{1/2}}{z - 1/24}\right] &&\\
&\qquad + \frac{1}{8} + \left[\frac{1}{2}(z - 1/6)\right]^{1/2} + \tfrac{1}{2}(z - 5/48)^{1/2}, && \frac{1}{6} < z \leqq \frac{2}{3},\\
&1, && z > \frac{2}{3}.
\end{aligned}
$$

4. Case C: $n = 3$. This is the first complicated case and reduces to the problem of evaluating the volume of the intersection of a sphere of radius

$$(z - 1/36)^{1/2},$$

with its center at $(\frac{1}{6}, \frac{1}{2}, \frac{5}{6})$, and a tetrahedron with vertices at $(0, 0, 0)$, $(0, 0, 1)$, $(0, 1, 1)$, and $(1, 1, 1)$. Whereas in the case $n = 2$ there are five intervals over which $F_2(z)$ is separately defined, when $n = 3$ there are eight: $(-\infty, 1/36)$,

TABLE I

Values of the CDF's of $n\omega_n^2$ for $n = 1, 2, 3$ and the asymptotic distribution at selected points

z	$F_1(z)$	$F_2(z)$	$F_3(z)$	$F(z)$
.11888	.37708	.46692	.47343	.50000
.14663	.50318	.57614	.57683	.60000
.16385	.56751	.63384	.63009	.65000
.18433	.63560	.68842	.68521	.70000
.20939	.71009	.73974	.74191	.75000
.24124	.79475	.79126	.79924	.80000
.28406	.89605	.84515	.85481	.85000
.34730	1.00000	.90296	.90617	.90000
.40520	1.00000	.94007	.93661	.93000
.46136	1.00000	.96554	.95723	.95000
.54885	1.00000	.98968	.97793	.97000
.74346	1.00000	1.00000	.99680	.99000
1.16786	1.00000	1.00000	1.00000	.99900

(1/36, 1/18), (1/18, 1/12), (1/12, 1/9), (1/9, 5/24), (5/24, 11/36), (11/36, 1) and (1, ∞). Partial results for these intervals are as follows, where $\zeta = z - 1/36$: $F_3^{(1)}(z) = 0$; $F_3^{(2)}(z) = 8\pi\zeta^{3/2}$; $F_3^{(3)}(z) = \frac{2}{3}\pi(3z - 1/9)$; $F_3^{(4)}(z) = \frac{2}{3}\pi(3z - 1/9) - 2\pi[4\zeta^{3/2} - 2^{\frac{1}{2}}\zeta + 2^{-\frac{1}{2}}/27] + 6\zeta^{3/2}\, V(1, \zeta^{-\frac{1}{2}}/6)$; $\cdots$; $F_3^{(8)}(z) = 1$, where $V(1, a)$ is the volume of the wedge-shaped segment of the sphere of unit radius, center at the origin, cut out by the two planes $x = a$ and $y = a$. It is possible to obtain expressions in closed form for $F_3(z)$ over all of the eight intervals; however their derivation is tedious and the expressions complicated.[1] A numerical evaluation was therefore undertaken by the RAND Numerical Analysis section. The result of these computations are shown in Table 1 along with the calculated values of $F_n(z)$ for $n = 1$ and 2, and for the asymptotic distribution. The values of $F_3(z)$ appear to be off by one in the fifth decimal place. The rapid convergence to the asymptotic distribution, especially in the more interesting region of the tail of distribution, seems clear

One other piece of evidence, although of a much weaker sort, is available that suggests that the asymptotic distribution is a good approximation to the exact distribution for small n. A sample of 400 values of $n\omega_n^2$ was produced for the case $n = 10$. Grouping into twenty cells using the 5th, 10th, 15th, $\cdots$, percentage points of the asymptotic distribution gave the following cell entries: 13, 19, 20, 18, 11, 21, 16, 18, 28, 17, 21, 22, 26, 18, 21, 16, 23, 25, 23, 24. Application of the χ^2 test gives a value of $\chi^2 = 17.5$. With 19 d.f. this value is exceeded with probability of approximately .55 Application of the Kolmogorov test statistic, Sup $| S_n(x) - F(x) |$, to the grouped data (for an approximate test) gives a value of 1 20. This value would be exceeded on the order of 11 per cent of the time under the null hypothesis.

Anderson and Darling in one of their papers [2] mention that "empirical study suggests that the asymptotic value is reached very rapidly, and it appears safe to use the asymptotic value for a sample size as large as 40." The results given above suggest the sample size for which it is reasonable to use the asymptotic distribution is likely to be more nearly 3 or 4, or perhaps 5.

For an allied form of the ω^2 test criterion, denoted by W_n^2 in [2] and formed by adding to (1) the weight function $\psi(X) = [F(X)(1 - F(X))]^{-1}$, an even more rapid convergence seems to occur. $F_1(z) = (1 - 4e^{-z-1})^{1/2}$ for the statistic W_1^2. Evaluating $F_1(z)$ at the 90, 95, and 99 asymptotic percentage points given in [2] yields .88716, .93292, and .98433 respectively.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193-212.

[2] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp 765-769.

[1] D. Anderson may publish the complete results later.

LIMITING DISTRIBUTIONS OF HOMOGENEOUS FUNCTIONS OF SAMPLE SPACINGS[1]

By Lionel Weiss

Cornell University

1. Summary. Suppose $T_1, T_2, \cdots, T_n$ are the lengths of n subintervals into which the interval [0, 1] is broken by $(n - 1)$ independent chance variables, each with a uniform distribution on [0, 1]. Moran [1], Kimball [2], and Darling [3] have shown that if r is a positive number, then the asymptotic distribution of $T_1^r + T_2^r + \cdots + T_n^r$ is normal. It is the purpose of this note to extend this result in two directions: more general functions of $T_1, \cdots, T_n$ are handled, and the joint distribution of several such functions is discussed. The proof is short and very simple.

2. Notation and assumptions. As already indicated, $T_1, T_2, \cdots, T_n$ are the n subintervals into which the unit interval is randomly broken. $U_1, U_2, \cdots, U_n$ are independent chance variables, each with the density function e^{-u} for $u \geqq 0$, zero for $u < 0$. $S_n = U_1 + U_2 + \cdots + U_n$. $V_i = U_i/S_n$ for $i = 1, \cdots, n$. It is known (and is very easily verified) that S_n is distributed independently of $(V_1, V_2, \cdots, V_n)$, and that the joint distribution of

$$(V_1, V_2, \cdots, V_n)$$

is exactly the same as the joint distribution of $T_1, T_2, \cdots, T_n$.

We are given k sequences of functions:

$$\{G_{1,n}(U_1, U_2, \cdots, U_n)\}, \cdots, \{G_{k,n}(U_1, U_2, \cdots, U_n)\},$$

$n = 1, 2, \cdots$. These functions are assumed to satisfy the following conditions:

(1) $G_{i,n}(U_1, \cdots, U_n)$ is homogeneous of order r_i for all n, r_i a positive quantity;

(2) the joint distribution of

$$\frac{G_{1,n}(U_1, \cdots, U_n) - A_1 n}{B_1\sqrt{n}}, \cdots, \frac{G_{k,n}(U_1, \cdots, U_n) - A_k n}{B_k\sqrt{n}}$$

approaches a k-variate normal distribution with zero means and covariance matrix C, say, as n increases. $A_1, \cdots, A_k$ and $B_1, \cdots, B_k$ are positive constants. (The results hold for any values of $A_1, \cdots, A_k$. The assumption that they are positive is merely a convenience.)

We denote the element of C in row i and column j by c_{ij}.

3. The asymptotic distribution of $G_{1,n}(T_1, \cdots, T_n), \cdots, G_{k,n}(T_1, \cdots, T_n)$. Theorem. *Under the assumptions of Sec. 2, the joint distribution of*

$$\frac{n^{r_1} G_{1,n}(T_1, \cdots, T_n) - A_1 n}{B_1\sqrt{n}}, \cdots, \frac{n^{r_k} G_{k,n}(T_1, \cdots, T_n) - A_k n}{B_k\sqrt{n}}$$

Received July 8, 1957.

[1] Research supported by the Office of Naval Research.

approaches a k-variate normal distribution with zero means and covariance matrix

$$\left\{c_{ij} - \frac{r_i r_j A_i A_j}{B_i B_j}\right\}$$

as n increases.

PROOF. By assumption, the distribution of the k-dimensional vector $\bar{V}(n)$ whose ith element is

$$\frac{G_{i,n}(U_1, \cdots, U_n) - A_i n}{B_i\sqrt{n}}$$

approaches the k-variate normal distribution with zero means and covariance matrix C. We rewrite the ith term of $\bar{V}(n)$ as

$$\frac{G_{i,n}(U_1, \cdots, U_n) - S_n^{r_i} A_i n^{1-r_i} + S_n^{r_i} A_i n^{1-r_i} - A_i n}{B_i\sqrt{n}}$$

Now S_n/n converges stochastically to one as n increases; therefore the distribution of the k-dimensional vector $\bar{V}'(n)$ whose ith element is

$$\frac{G_{i,n}(U_1, \cdots, U_n) - S_n^{r_i} A_i n^{1-r_i} + S_n^{r_i} A_i n^{1-r_i} - A_i n}{\left(\frac{S_n}{n}\right)^{r_i} B_i\sqrt{n}}$$

approaches the k-variate normal distribution with zero means and covariance matrix C. $\bar{V}'(n)$ may be written as the sum of two vectors, $\bar{V}_1(n)$ and $\bar{V}_2(n)$, whose ith elements are respectively

$$\frac{n^{r_i} G_{i,n}(V_1, \cdots, V_n) - A_i n}{B_i\sqrt{n}}$$

and

$$\frac{A_i n - n^{r_i+1} A_i S_n^{-r_i}}{B_i\sqrt{n}}$$

We note that $\bar{V}_1(n)$ and $\bar{V}_2(n)$ are distributed independently of each other.

Next we examine the distribution function, say $F_n(x_1, \cdots, x_k)$, of $\bar{V}_2(n)$.

$$F_n(x_1, \cdots, x_k) = \Pr\left[\frac{A_i n - n^{r_i+1} A_i S_n^{-r_i}}{B_i\sqrt{n}} \leq x_i;\ i = 1, \cdots, k\right]$$

$$= \Pr\left[\frac{S_n - n}{\sqrt{n}} \leq \sqrt{n}\left\{\left(\frac{A_i n}{A_i n - \sqrt{n}\, B_i x_i}\right)^{\frac{1}{r_i}} - 1\right\},\right.$$

As n increases, the distribution of $(S_n - n)/\sqrt{n}$ approaches the standard normal distribution, by the univariate central-limit theorem. And for any fixed x_i,

$$\sqrt{n}\left\{\left(\frac{A_i n}{A_i n - \sqrt{n}\, B_i x_i}\right)^{\frac{1}{r_i}} - 1\right\} \to \frac{B_i x_i}{r_i A_i}$$

as n increases. Thus, if Z denotes a chance variable with a standard normal distribution, $F_n(x_1, \cdots, x_k)$ approaches

$$\Pr\left[\frac{r_i A_i Z}{B_i} \leqq x_i\, ; i = 1, \cdots, k\right]$$

for each vector $(x_1, \cdots, x_k)$.

Next, we denote by $\rho_{1,n}(t_1, \cdots, t_k)$ the characteristic function of $\bar{V}_1(n)$, by $\rho_{2,n}(t_1, \cdots, t_k)$ the characteristic function of $\bar{V}_2(n)$, and by $\rho_n(t_1, \cdots, t_k)$ the characteristic function of $\bar{V}'(n)$.

We have $\rho_n(t_1, \cdots, t_k) = \rho_{1,n}(t_1, \cdots, t_k) \cdot \rho_{2,n}(t_1, \cdots, t_k)$, or

$$\rho_{1,n}(t_1, \cdots, t_k) = \frac{\rho_n(t_1, \cdots, t_k)}{\rho_{2,n}(t_1, \cdots t_k)}.$$

As n increases,

$$\rho_n(t_1, \cdots, t_k) \to \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{k}\sum c_{ij} t_i t_j\right\}$$

and

$$\rho_{2,n}(t_1, \cdots, t_k) \to \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{k} \frac{t_j r_j A_j}{B_j}\right]^2\right\}.$$

Therefore, as n increases,

$$\rho_{1,n}(t_i, \cdot \cdot, t_k) \to \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{k}\sum t_i t_j \left[c_{ij} - \frac{r_i r_j A_i A_j}{B_i B_j}\right]\right\}.$$

This proves the theorem.

REFERENCES

[1] P. Moran, "The random division of an interval," *J. Roy. Stat. Soc., Suppl.*, Vol. 9 (1947), pp. 92–98.

[2] B. F. Kimball, "Some basic theorems for developing tests of fit for the case of non-parametric probability distribution functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 540–548.

[3] D. A. Darling, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239–253.

ANOTHER COUNTABLE MARKOV PROCESS WITH ONLY INSTANTANEOUS STATES

By David Blackwell[1]

University of California, Berkeley

Let P be the transition function for a Markov process with a countable state space A and stationary transition probabilities, i.e., P is a nonnegative function defined for all triples (a, b, t) with $a \in A$, $b \in A$, and t a nonnegative real number, satisfying

$$(1) \qquad P(a, b, 0) = 1 \text{ if } a = b, \qquad 0 \text{ if } a \neq b.$$

$$(2) \qquad \sum_b P(a, b, t) = 1 \quad \text{for all} \quad a, t,$$

and

$$(3) \qquad P(a, b, s + t) = \sum_{c \in A} P(a, c, s)P(c, b, t) \quad \text{for all} \quad s \geqq 0, \qquad t \geqq 0, a, b.$$

We shall suppose, as usual, that P is continuous at $t = 0$, i.e.,

$$(4) \qquad P(a, a, t) \to 1 \text{ as } t \to 0 \text{ for all } a.$$

It is well known that, for any P satisfying (1), (2), (3), and (4), $P'(a, a, 0)$ exists for all a (it may be negatively infinite). Following P. Lévy [2], a state is called "instantaneous" if $P'(a, a, 0) = -\infty$. Examples of processes with all states instantaneous have been given by Feller and McKean [2] and by Dobrushin [1]. The purpose of this note is to describe a third example, somewhat simpler than those previously given.

We first describe the process informally, after which we define P and verify (1), (2), (3), and (4) and $P'(a, a, 0) = -\infty$ for all a directly. Let $X_1(t)$, $X_2(t)$, $\cdots$ be a sequence of Markov processes, independent of each other, each with two states 0 and 1. We suppose $X_n(0) = 0$ for all n. Let $X_n(t)$ be characterized by the parameters λ_n, μ_n:

$$\Pr\{X_n(t + h) = 1 \mid X_n(t) = 0\} = \lambda_n h + o(h),$$

$$\Pr\{X_n(t + h) = 0 \mid X_n(t) = 1\} = \mu_n h + o(h).$$

Our process $X(t)$ will be the joint process $X_1(t)$, $X_2(t)$, $\cdots$ which is clearly a Markov process. To insure that $X(t)$ has only a countable set of states, we

Received July 11, 1957.

[1] This paper was supported in part by funds provided under Contract AF-41(657)-29 with the Air Research and Development Command, USAF School of Aviation Medicine, Randolph Field, Texas.

determine λ_n, μ_n so that, at each time t, with probability 1, $X_n(t) = 0$ for al but a finite number of n. Since

$$\Pr(X_n(t) = 0 \mid X_n(0) = 0) = \frac{\mu_n}{\mu_n + \lambda_n} + \frac{\lambda_n}{\mu_n + \lambda_n} e^{-(\lambda_n+\mu_n)t}$$

$$\geqq \frac{\mu_n}{\mu_n + \lambda_n},$$

this will occur if

$$\prod_n \frac{\mu_n}{\mu_n + \lambda_n} > 0, \tag{5}$$

i.e.,

$$\sum_n \frac{\lambda_n}{\lambda_n + \mu_n} < \infty.$$

A state is instantaneous if and only if the probability of remaining ii it throughout an interval is zero. Since the probability that $X_n(t) = 0$ through out T, $T + h$ given that $X_n(T) = 0$ is $e^{-\lambda_n h}$, the chance that the state $X(T$ with $X_n(T) = 0$ for $n \geqq N$ will persist throughout T, $T + h$ is at most

$$\prod_N^\infty e^{-\lambda_n h} = e^{-h(\lambda_N + \lambda_{N+1} + \cdots)},$$

and will be zero if

$$\sum_n \lambda_n = \infty. \tag{6}$$

Thus any choice of $\{\lambda_n\}$, $\{\mu_n\}$ satisfying (5) and (6) yields an example of a proc ess with only instantaneous states.

Formally, the set A of states is the set of all infinite sequences

$$a = (\epsilon_1, \epsilon_2, \cdots)$$

of 0's and 1's with only finitely many 1's. Let $\{\lambda_n\}$, $\{\mu_n\}$ be sequences of posi tive numbers satisfying (5) and (6), let

$$R_n(0, 0, t) = \frac{\mu_n}{\mu_n + \lambda_n} + \frac{\lambda_n}{\mu_n + \lambda_n} e^{-(\lambda_n+\mu_n)t},$$

$$R_n(1, 1, t) = \frac{\lambda_n}{\mu_n + \lambda_n} + \frac{\mu_n}{\mu_n + \lambda_n} e^{-(\lambda_n+\mu_n)t},$$

$$R_n(0, 1, t) = 1 - R_n(0, 0, t),$$

$$R_n(1, 0, t) = 1 - R_n(1, 1, t),$$

and define, for any two states $a = (\epsilon_1, \epsilon_2, \cdots)$ and $b = (\delta_1, \delta_2, \cdots)$ and any $t \geqq 0$,

$$(7) \qquad P(a, b, t) = \prod_{n=1}^{\infty} R_n(\epsilon_n, \delta_n, t).$$

Denote by A_N the set of all states $a = (\epsilon_1, \epsilon_2, \cdots)$ with $\epsilon_n = 0$ for all $n > N$. For $a \in A_N$ and any $M \geqq N$, we have

$$\sum_{b \in A_M} P(a, b, t) = h_M(t) \sum_{\delta_1, \cdots, \delta_M} \prod_1^M R_n(\epsilon_n, \delta_n, t)$$

$$= h_M(t) \prod_1^M (R_n(\epsilon_n, 0, t) + R_n(\epsilon_n, 1, t)) = h_M(t),$$

where

$$(8) \qquad h_M(t) = \prod_{M+1}^{\infty} R_n(0, 0, t) \geqq \prod_{M+1}^{\infty} \frac{\mu_n}{\mu_n + \lambda_n} = V_M$$

From (8), $h_M(t) \to 1$ as $M \to \infty$, so that (2) is verified. For (3), say $a \in A_N$, $b \in A_N$. For $M \geqq N$,

$$\sum_{c \in A_M} P(a, c, s)P(c, b, t)$$

$$= h_M(s)h_M(t) \sum_{\alpha_1, \cdots, \alpha_M} \prod_{n=1}^{M} R_n(\epsilon_n, \alpha_n, s)R_n(\alpha_n, \delta_n, t)$$

$$= h_M(s)h_M(t) \prod_{n=1}^{M} \left(\sum_{\alpha=0}^{1} R_n(\epsilon_n, \alpha, s)R_n(\alpha, \delta_n, t)\right)$$

$$= h_M(s)h_M(t) \prod_{n=1}^{M} R_n(\epsilon_n, \delta_n, s + t) \to P(a, b, s + t) \quad \text{as} \quad M \to \infty.$$

For (4), if $a \in A_N$ and $M \geqq N$,

$$P(a, a, t) \geqq \left(\prod_1^M R_n(\epsilon_n, \epsilon_n, t)\right) V_M,$$

so that

$$\liminf_{t \to 0} p(a,a,t) \geqq V_M.$$

Since this holds for all M and $V_M \to 1$ as $M \to \infty$, (4) is verified. Finally, since, for $a \in A_N$ and $M \geqq N$ we have, for all $k \geqq 1$

$$P(a, a, t) \leqq h_{M,k}(t) = \prod_{M+1}^{M+k} R_n(0, 0, t),$$

and since $P(a, a, 0) = h_{M,k}(0) = 1$,

$$P'(a, a, 0) \leqq h'_{M,k}(0) = -\sum_{M+1}^{M+k} \lambda_n,$$

so that (6) implies $P'(a, a, 0) = -\infty$.

REFERENCES

[1] R. L. DOBRUSHIN, "An example of a countable homogeneous Markov process all states of which are instantaneous", (Russian) *Teor. Veroyatnost. i Primenen.* 1(1956), 481–485.

[2] WILLIAM FELLER AND HENRY P. MCKEAN, JR., "A diffusion equivalent to a countable Markov chain," *Proceedings Nat. Acad. of Sci.*, Vol. 42, No. 6 (1956), pp. 351-354.

[3] P. LÉVY, "Systems Markoviens et stationnaires. Cas denombrable," *Annales Sci. de l'Ecole Normale Superieure de Paris*, Vol. 68 (1951), pp. 327–381.

SPACINGS GENERATED BY MIXED SAMPLES

BY LIONEL WEISS[1]

Cornell University

1. Summary and introduction. Suppose $X(1, 1), X(1, 2), \cdots, X(1, n_1), X(2, 1), \cdots, X(2, n_2), \cdots, X(k, 1), \cdots, X(k, n_k)$ are independent chance variables, $X(i, j)$ having the probability density function $f_i(x)$, for $j = 1, \cdots, n_i$, $i = 1, \cdots, k$. We assume that for each i, $f_i(x)$ is bounded and has at most a finite number of discontinuities. We denote $n_1 + n_2 + \cdots + n_k$ by N, and we assume that n_i/N is equal to r_i, where r_i is a given positive number. Let $Y_1 \leqq Y_2 \leqq \cdots \leqq Y_N$ denote the ordered values of the N observations

$$X(1, 1), \cdots, X(k, n_k).$$

Define W_i as $Y_{i+1} - Y_i$ for $i = 1, \cdots, N - 1$. For any given nonnegative t, let $R_N(t)$ denote the proportion of the values $W_1, \cdots, W_{N-1}$ which are greater than t/N. Let $S(t)$ denote

$$\int_{-\infty}^{\infty} (r_1f_1(x) + r_2f_2(x) + \cdots + r_kf_k(x)) \exp\{-t[r_1f_1(x) + \cdots + r_kf_k(x)]\}\, dx$$

and $V(N)$ denote $\sup_{t\geqq 0} | R_N(t) - S(t) |$. Then it is shown that $V(N)$ converges stochastically to zero as N increases. This is a generalization of [1], where k was equal to unity. The result is applied to find the asymptotic behavior of ranks in a k-sample problem.

2. Proof of the stochastic convergence of $V(N)$. As in [1], if it can be shown that $R_N(t)$ converges stochastically to $S(t)$ for each positive t, the convergence of $V(N)$ follows. Therefore we fix a positive value for t.

We define the chance variable $Z(i, j, N)$ to be equal to unity if no observations fall in the half-open interval $[(X(i, j), X(i, j) + t/N]$, and equal to zero otherwise. We denote $1/N \sum_{i=1}^{k} \sum_{j=1}^{n_i} Z(i, j, N)$ by $K(N)$. Clearly,

$$K(N) = (1 - 1/N)R_N(t) + 1/N,$$

Received July 17, 1957; revised August 26, 1957.

[1] Research sponsored by the Office of Naval Research.

so our purpose is accomplished if we show that $K(N)$ converges stochastically to $S(t)$ as N increases.

We denote $\int_{-\infty}^{x} f_i(x)\,dx$ by $F_i(x)$.

$$E\{Z(i, j, N)\} = \int_{-\infty}^{\infty} \left[1 - F_i\left(x + \frac{t}{N}\right) + F_i(x)\right]^{n_i - 1} \cdot \prod_{h \neq i} \left[1 - F_h\left(x + \frac{t}{N}\right) + F_h(x)\right]^{n_h} dF_i(x).$$

But with the exception of a finite number of points, $F_i(x + t/N) - F_i(x)$ can be written as $[f_i(x) + \epsilon_i(x, t/N)]t/N$, where $\epsilon_i(x, t/N)$ approaches zero as N increases, for each x. Since $f_i(x)$ is bounded $(i = 1, \cdots, k)$, it follows easily that $E\{Z(i, j, N)\}$ approaches

$$\int_{-\infty}^{\infty} \exp\{-t[r_1 f_1(x) + \cdots + r_k f_k(x)]\}\, dF_i(x)$$

as N increases. It follows immediately that $E\{K(N)\}$ approaches $S(t)$ as N increases.

Next we examine variance $\{K(N)\}$, which equals $N^{-2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}$ variance $\{Z(i, j, N)\} + 1/N^2 \sum\sum_{(i,j) \neq (g,h)}\sum\sum \operatorname{cov}\{Z(i, j, N), Z(g, h, N)\}$. The first term in this last expression clearly approaches zero as N increases, since there are N uniformly bounded terms in the sum. We shall show that the second term also approaches zero by showing that the covariances approach zero uniformly. Since there are $N(N-1)$ covariances, the factor $1/N^2$ guarantees the approach to zero. If $i \neq g$, $E\{Z(i, j, N) \cdot Z(g, h, N)\}$ is equal to

$$\iint_{|x-y| > \frac{t}{N}} \prod_{b \neq i,g} \left[1 - F_b\left(x + \frac{t}{N}\right) + F_b(x) - F_b\left(y + \frac{t}{N}\right) + F_b(y)\right]^{n_b}$$
$$\cdot \left[1 - F_i\left(x + \frac{t}{N}\right) + F_i(x) - F_i\left(y + \frac{t}{N}\right) + F_i(y)\right]^{n_i - 1}$$
$$\cdot \left[1 - F_g\left(x + \frac{t}{N}\right) + F_g(x) - F_g\left(y + \frac{t}{N}\right) + F_g(y)\right]^{n_g - 1} dF_i(x)\, dF_g(y).$$

By computations similar to those used on $E\{Z(i, j, N)\}$, it follows that

$$E\{Z(i, j, N)\ Z(g, h, N)\}$$

approaches

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-t[r_1 f_1(x) + \cdots + r_k f_k(x)]\} \cdot \exp\{-t[r_1 f_1(y) + \cdots + r_k f_k(y)]\} \cdot dF_i(x)\, dF_g(y)$$

and from this it follows that $\operatorname{cov}\{Z(i, j, N), Z(g, h, N)\}$ approaches zero as N increases. In the same way, it follows that

$$\operatorname{cov}\{Z(i, j, N), Z(i, h, N)\}$$

approaches zero ($j \neq h$). Thus variance$\{K(N)\}$ approaches zero as N increases so $K(N)$ converges stochastically to $S(t)$, as does $R_N(t)$. Therefore we have shown that $V(N)$ converges stochastically to zero as N increases.

3. Application to ranks in k-samples. Define $T(i, j)$ as $F_1(X(i, j))$. Then $T(1, 1), \cdots, T(1, n_1)$ have unform distributions. Let $G_i(x)$ denote the resulting distribution function for $T(i, j)$. We assume that $G_i(x)$ allows a density function $g_i(x)$ (then $g_i(x)$ is zero outside the interval [0, 1], is bounded, and has a finite number of discontinuities). Let $V_1 \leqq V_2 \leqq \cdots \leqq V_{N-n_1}$ denote the ordered values of $T(2, 1), \cdots, T(k, n_k)$, and let V_0 equal zero, V_{N-n_1+1} equal one. Let S_i denote the number of $T(1, j)$'s which lie in the interval

$$[V_{i-1}, V_i], \qquad i = 1, \cdots, N - n_1 + 1.$$

For each nonnegative integer r, let $Q_n(r)$ be the proportion of values among $S_1, \cdots, S_{N-n_1+1}$ which are equal to r. Define $g(y)$ as $\sum_{i=2}^{k} (r_i/(1 - r_1))g_i(y)$, and α as $(r_1/(1 - r_1))$. Define $Q(r)$ as

$$\alpha^r \int_0^1 \frac{g^2(y)}{[\alpha + g(y)]^{r+1}} \, dy.$$

Then it follows from the results above, using also the argument in [2], that $\sup_{r \geqq 0} | Q_N(r) - Q(r) |$ converges stochastically to zero as N increases. This can be used to show that certain tests of the hypothesis

$$F_1(x) = F_2(x) =. \cdots = F_k(x)$$

are consistent. The discussion parallels that found in [2].

REFERENCES

[1] L. Weiss, "The stochastic convergence of a function of sample successive differences," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 532–536.

[2] J. R. Blum and L. Weiss, "Consistency of certain two-sample tests," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 242–246.

CORRECTION TO "AN EXTENSION OF THE KOLMOGOROV DISTRIBUTION"

By Jerome Blackman

Syracuse and Cornell Universities

1. Summary. It has been pointed out by J. H. B. Kemperman that an error in [1] invalidates the formulas arrived at in that paper. It is the purpose of this note to supply the correct formulas for the probabilities of Theorems 1 and 2. An Appendix by Professor Kemperman is included.

Received February 18, 1957; revised August 14, 1957.

approaches zero ($j \neq h$). Thus variance$\{K(N)\}$ approaches zero as N increa
so $K(N)$ converges stochastically to $S(t)$, as does $R_N(t)$. Therefore we h
shown that $V(N)$ converges stochastically to zero as N increases.

3. Application to ranks in k-samples. Define $T(i, j)$ as $F_1(X(i, j))$. T
$T(1, 1), \cdots, T(1, n_1)$ have unform distributions. Let $G_i(x)$ denote the res
ing distribution function for $T(i, j)$. We assume that $G_i(x)$ allows a den
function $g_i(x)$ (then $g_i(x)$ is zero outside the interval $[0, 1]$, is bounded,
has a finite number of discontinuities). Let $V_1 \leqq V_2 \leqq \cdots \leqq V_{N-n_1}$ de
the ordered values of $T(2, 1), \cdots, T(k, n_k)$, and let V_0 equal zero, V_{N-}
equal one. Let S_i denote the number of $T(1, j)$'s which lie in the interval

$$[V_{i-1}, V_i], \qquad i = 1, \cdots, N - n_1 + 1.$$

For each nonnegative integer r, let $Q_n(r)$ be the proportion of values am
$S_1, \cdots, S_{N-n_1+1}$ which are equal to r. Define $g(y)$ as $\sum_{i=2}^{k} (r_i/(1 - r_1))g_i(y)$,
α as $(r_1/(1 - r_1))$. Define $Q(r)$ as

$$\alpha^r \int_0^1 \frac{g^2(y)}{[\alpha + g(y)]^{r+1}} \, dy.$$

Then it follows from the results above, using also the argument in [2], t
$\sup_{r \geqq 0} | Q_N(r) - Q(r) |$ converges stochastically to zero as N increases. T
can be used to show that certain tests of the hypothesis

$$F_1(x) = F_2(x) = \cdots = F_k(x)$$

are consistent. The discussion parallels that found in [2].

REFERENCES

[1] L. Weiss, "The stochastic convergence of a function of sample successive differenc
Ann. Math. Stat., Vol. 26 (1955), pp. 532–536.
[2] J. R. Blum and L. Weiss, "Consistency of certain two-sample tests," *Ann. Math. S*
Vol. 28 (1957), pp. 242–246.

CORRECTION TO "AN EXTENSION OF THE KOLMOGOROV DISTRIBUTION"

By Jerome Blackman

Syracuse and Cornell Universities

1. Summary. It has been pointed out by J. H. B. Kemperman that an er
in [1] invalidates the formulas arrived at in that paper. It is the purpose of t
note to supply the correct formulas for the probabilities of Theorems 1 and
An Appendix by Professor Kemperman is included.

Received February 18, 1957; revised August 14, 1957.

2. Introduction. The major error in [1] lies in the mappings on pp. 3[illegible] Corrections can be made for this but unfortunately the resulting formulas more complicated than before. A smaller error appears in the statement $N(A_{2i}) = N(B_{2i})$, but this is easily corrected. The new formulas are so m[illegible] more complicated that it has not seemed worthwhile to correct Corollari[illegible] and 2 which are hereby retracted. The corrected statements of the main th[illegible] follow.

THEOREM 1. *Let* $x_1, x_2, \cdots, x_n, x_1', x_2', \cdots, x_{nk}'$ *be a sequence of* $n(1+k)$ *independent random variables with a common continuous distribution* $F(x)$ [illegible] $F_n(x)$ *and* $G_{nk}(x)$ *be empiric distributions based on the first* n *and second* kn [illegible] *dom variables respectively. Then*

$$P(-y < G_{nk}(s) - F_n(s) < x \quad \text{for all} \quad s)$$

$$= 1 - \binom{(k+1)n}{n}^{-1} \sum_{i=1}^{\infty} \{N(A_{2i-1}) + N(B_{2i-1}) - N(A_{2i}) - N(B \ldots$$

where the N *functions are given in* (1), (2), *and* (3).

THEOREM 2.

$$P(-y < F(s) - F_n(s) < x \quad \text{for all} \quad s)$$

$$= 1 - \sum_{i=1}^{\infty} \{S(A_{2i-1}) + S(B_{2i-1}) - S(A_{2i}) - S(B \ldots$$

where the N *functions are given in* (5), (6), *and* (7).

3. Corrections. The point of departure [illegible] a formula for $N(\mathfrak{a}_0)$ is given. It is [illegible] by the total number of paths $\binom{(k+1)n}{n}$ [illegible] the analytical expressions for the N functions [illegible] of the A_i and B_i classes described at the [illegible] clusions drawn there about the mappings are [illegible] consider the image of a path from A_i [illegible] path which starts from the origin [illegible] stops at least once at the point a. The [illegible] with the set of paths which start [illegible] stop at least once at a. Because the [illegible] path which reaches $2a + b$ will at [illegible] images of the A_i and B_i under the mappings [illegible] the points which the path must reach [illegible] which the path must stop, in order [illegible] 2. In all cases the mapping is 1:1 [illegible] set of paths which reach the point [illegible] quently stop in order at all the points [illegible]

TABLE 1

A_{2i-1}	$i(\alpha + \beta) - \beta$	$(i-1)(\alpha+\beta) - \beta, \quad (i-2)(\alpha+\beta) - \beta, \cdots \alpha$
A_{2i}	$i(\alpha + \beta)$	$(i-1)(\alpha+\beta), \quad (i-2)(\alpha+\beta), \cdots (\alpha+\beta)$
B_{2i-1}	$i(\alpha + \beta) - \alpha$	$(i-1)(\alpha+\beta) - \alpha, \quad (i-2)(\alpha+\beta) - \alpha, \cdots \beta$
B_{2i}	$i(\alpha + \beta)$	$i(\alpha+\beta) - \beta, \quad (i-1)(\alpha+\beta) - \beta, \cdots \alpha$

As a preliminary step consider the number of ways a path consisting of i steps to the left and $ki - \alpha$ steps to the right can go from α to 0 without touching α after the first step. Let the number of these paths be $H_\alpha(i)$. While this number can be computed by elementary methods a more elegant formula has been obtained by Professor Kemperman, namely,

$$H_\alpha(i) = (k+1) \sum_{0 \le r < \alpha/(k+1)} \frac{(-1)^r}{(i-r)(k+1)-1} \binom{(i-r)(k+1)-1}{i-r} \binom{\alpha - 1 - kr}{r} - \frac{\alpha}{(k+1)i - \alpha} \binom{(k+1)i - \alpha}{i}. \tag{1}$$

The proof of this is contained in the appendix.

The number of ways of going from 0 to α after exactly j steps to the left and $kj + \alpha$ steps to the right will be indicated by $J(\alpha, j)$ where

$$J(\alpha, j) = \binom{(k+1)j + \alpha}{j}. \tag{2}$$

Combining the results of Table 1 and the definitions of H and J we see that

$$\begin{aligned} N(A_{2i-1}) &= \sum_{j_1 + \cdots + j_{i+1} = n} J(i(\alpha+\beta) - \beta, j_1) \prod_{k=2}^{i} H_{\alpha+\beta}(j_k) H_\alpha(j_{i+1}), \\ N(A_{2i}) &= \sum_{j_1 + \cdots + j_{i+1} = n} J(i(\alpha+\beta), j_1) \prod_{k=2}^{i+1} H_{\alpha+\beta}(j_k), \\ N(B_{2i-1}) &= \sum_{j_1 + \cdots + j_{i+1} = n} J(i(\alpha+\beta) - \alpha, j_1) \prod_{k=2}^{i} H_{\alpha+\beta}(j_k) H_\beta(j_{i+1}), \\ N(B_{2i}) &= \sum_{j_1 + \cdots + j_{i+2} = n} J(i(\alpha+\beta), j_1) H_\beta(j_2) \prod_{k=3}^{i+1} H_{\alpha+\beta}(j_k) H_\alpha(j_{i+2}). \end{aligned} \tag{3}$$

This completes Theorem 1. The infinite series occurring in this theorem is really a finite series in view of $N(A_{2i-1}) = \cdots = N(B_{2i}) = 0$ For

$$i > nk/(\alpha + \beta).$$

To get Theorem 2 it is only necessary to take the limit as $k \to \infty$ in the various formulas given above. By Stirling's formula,

$$(4) \qquad \binom{t(1+k)}{t} = \frac{t^t}{t!}O((1+k)^t) \quad \text{as} \quad k \to \infty.$$

Here and below we will use $a_k = 0(b_k)$ to mean $\lim_{k\to\infty} a_k/b_k = 1$.

Using (4) and a few more applications of Stirling's formula and remembering that $\alpha = -[-xkn]$ and $\beta = -[-ykn]$, we obtain

$$(5) \qquad H_\alpha(i) = \left\{ \sum_{0\leq r<xn} (-1)^r \frac{(i-r)^{i-r-1}(xn-r)^r}{(i-r)!r!} - xn(i-xn)^{i-1}/i! \right\} \cdot O((1+k)^i) = \bar{H}_x(i)O((1+k)^i)$$

where the last equality defines $\bar{H}_x(i)$ Using (4) again

$$(6) \qquad J(\alpha, j) = \frac{1}{j!}(j+xn)^j O((1+k)^j) = \bar{J}(x, j)O((1+k)^j)$$

and

$$\binom{(k+1)n}{n}^{-1} = \frac{n!}{n^n}O((1+k)^{-n})$$

Combining these results and (3) the following equations are obtained:

$$(7) \qquad \begin{aligned}
&\lim_{k\to\infty} \binom{(k+1)n}{n}^{-1} N(A_{2i-1}) \\
&\qquad = \sum_{j_1+\cdots+j_{i+1}=n} \bar{J}(i(x+y)-y, j_1) \prod_{k=2}^{i} \bar{H}_{x+y}(j_k)\bar{H}(j_{i+1}) \\
&\qquad = \bar{N}(A_{2i-1}), \\
&\lim_{k\to\infty} \binom{(k+1)n}{n}^{-1} N(A_{2i}) \\
&\qquad = \sum_{j_1+\cdots+j_{i+1}=n} \bar{J}(i(x+y), j_1) \prod_{k=2}^{i+1} \bar{H}_{x+y}(j_k) \\
&\qquad = \bar{N}(A_{2i}), \\
&\lim_{k\to\infty} \binom{(k+1)n}{n}^{-1} N(B_{2i-1}) \\
&\qquad = \sum_{j_1+\cdots+j_{i+2}=n} \bar{J}(i(x+y)-x, j_1) \prod_{k=2}^{i} \bar{H}_{x+y}(j_k)\bar{H}_y(j_{i+1}) \\
&\qquad = \bar{N}(B_{2i-1}), \\
&\lim_{k\to\infty} \binom{(k+\ )n}{n}^{-1} N(B_{2i}) \\
&\qquad = \prod_{j_1+\cdots+j_{i+2}=n} \bar{J}(i(x+y), j_1)\bar{H}_y(j_2) \prod_{k=3}^{i+1} \bar{H}_{x+y}(j_k)\bar{H}(j_{i+2}) \\
&\qquad = \bar{N}(B_{2i}).
\end{aligned}$$

This completes Theorem 2

Attention should be drawn to a paper of Korolyuk [2] wherein the author gives different versions of the probabilities we have presented for the case $x = y$.

REFERENCES

[1] J. BLACKMAN, "An extension of the Kolmogorov distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 513–520.

[2] V. S. KOROLYUK, "On the difference of the empirical distribution of two independent samples," *Izv. Akad. Nauk. SSSR*, Vol. 19 (1955), pp. 81–96.

APPENDIX

BY J. H. B. KEMPERMAN

By a *path* of length n we shall mean an ordered sequence of $n + 1$ integers $(z_0, \cdots, z_n)$, such that

$$z_i - z_{i-1} \geqq -1 \qquad (i = 1, \cdots, n).$$

For each path $\pi_n = (z_0, \cdots, z_n)$, let

$$P(\pi_n) = \prod_{i=1}^{n} p(z_i - z_{i-1}),$$

(the weight or "probability" of π_n). Here, the $p_i = p(i)$, $(i = -1, 0, +1, \cdots)$, denote given (real or complex) numbers, $p(-1) \neq 0$. Finally, let

$$e_z(n) = \sum_{\pi_n}{}' p(\pi_n),$$

the summation being extended over all the paths $\pi_n = (z_0, \cdots, z_n)$ with $z_0 = 0$, $z_n = z$, $z_i \neq z (i = 0, 1, \cdots, n - 1)$.

THEOREM. *For* $n = 1, 2, \cdots$,

$$(8) \qquad e_z(n) = -zr_z(n)/n + \sum_{j=1}^{\infty} j(j+1)p_j \sum_{0<m\leqq+z} r_z(-m)r_{-j}(m+n-1)/(m+n-1).$$

Here, for arbitrary integers h *and* s, $r_h(s)$ *is defined as the coefficient of* w^{h+s} *in the formal development*

$$(p_{-1} + p_0 w + p_1 w^2 + \cdots)^s = \sum_h r_h(s) w^{h+s};$$

especially, $r_h(s) = 0$ *if* $h + s < 0$.

PROOF. Let n and z be given integers, $n \geqq 1$. For any path $(z_0, \cdots, z_n)$ with $z_0 = 0$, $z_n = z$, we have

$$z_i - z_{i-1} = z - \sum_{\substack{\nu=1 \\ \nu\neq i}}^{n} (z_\nu - z_{\nu-1}) \leqq z + n - 1,$$

$(i = 1, \cdots, n-1)$, thus, $c_z(n)$ does not depend on the p_i with $i \geqq n + z$. Further, $r_h(s)$ does not depend on the p_i with $i \geqq h + s$, hence, the inner sum in (8) does not depend on the p_i with $i \geqq n + z$; moreover, the jth inner sum equals 0 when $j \geqq n + z$. Consequently, it suffices to prove the theorem for the special case that $p_i = 0$ for i sufficiently large.

In this case,

$$f(w) = \sum_{i=-1}^{\infty} p_i w^i$$

is analytic at each point $w \neq 0$. Further, for $|w|$ sufficiently small

$$f(w)^s = \sum_{-\infty}^{\infty} r_h(s) w^h, \tag{9}$$

hence, for $s \geqq 0$

$$r_h(s) = \sum_{\pi_s}{}' P(\pi_s),$$

summing over all the paths $\pi_s = (z_0, \cdots, z_s)$ with $z_0 = 0$, $z_s = h$. Observing that to each path $(z_0, \cdots, z_n)$ with $z_n = z$ there corresponds a unique integer m with $0 \leqq m \leqq n$, $z_i \neq z (i = 0, 1, \cdots, m-1)$, $z_m = z$, it follows that

$$r_z(n) = \sum_{m=0}^{n} c_z(m) r_0(n-m) \qquad (n = 0, 1, \cdots),$$

hence,

$$E_z = R_z / R_0, \tag{10}$$

where

$$R_h = \sum_{n=0}^{\infty} r_h(n) t^n, \qquad E_z = \sum_{n=0}^{\infty} c_z(n) t^n, \tag{11}$$

t denoting a sufficiently small parameter, $t \neq 0$.

Further, from (9), for each integer h,

$$R_h + \sum_{-h \leqq n < 0} r_h(n) t^n = \sum_{n=-h}^{\infty} r_h(n) t^n$$

$$= \sum_{n=-h}^{\infty} \frac{t^n}{2\pi\sqrt{-1}} \int_{|w|=R} f(w)^n w^{-h-1}\, dw = \frac{1}{2\pi\sqrt{-1}} \int_{|w|=R} \frac{(wf(w)t)^{-h}}{w - twf(w)}\, dw,$$

where R denotes a fixed positive number with $f(w) \neq 0$ for $0 < |w| \leqq R$. Here, from $p(-1) \neq 0$, the integrand is regular at $w = 0$. Moreover, for $t \neq 0$, $|t|$ sufficiently small, the equation $f(\xi) = t^{-1}$ has a unique solution satisfying $0 < |\xi| < R$. Thus,

$$R_h = (-t\xi f'(\xi))^{-1} \xi^{-h} - \sum_{0 < m \leqq h} r_h(-m) t^{-m}.$$

Finally, (10) and

$$\xi f'(\xi) = \xi f'(\xi) + f(\xi) - t^{-1} = -t^{-1} + \sum_{j=0}^{\infty} (j+1)p_j\xi^j$$

imply

$$E_z = \xi^{-z} + (-1 + t\sum_{j=0}^{\infty}(j+1)p_j\xi^j) \sum_{0<m\leqq z} r_2(-m)t^{-m}.$$

In view of (11), it suffices to prove that, for each integer h and $|t|$ sufficiently small, $t \neq 0$,

$$\xi^{-h} = -h \sum_{\substack{m=-h \\ m\neq 0}}^{\infty} r_h(m)t^m/m + c_h,$$

where c_h denotes a constant. Now, for $|t|$, $|\xi|$ small, the mapping $t \to \xi$ defined by $f(\xi) = t^{-1}$ is a 1:1 analytic transformation. Hence, integrating along a small positively oriented circle about 0, we have, for $m \neq 0$,

$$\int \xi^{-h}t^{-m-1}\,dt = -\int \xi^{-h}\,d(f(\xi)^m/m) = -\frac{h}{m}\int f(\xi)^m\xi^{-h-1}\,d\xi = -2\pi\sqrt{-1}\,\frac{h}{m}r_h(m).$$

REMARK. Results and methods analogous to the above may be found in the paper "The passage problem for a stationary Markov chain" by J. H. B. Kemperman, to appear in these Annals.

Let k be a fixed positive integer and choose $p(-1) = p(k) = 1$, $p(i) = 0$ for $i \neq -1, k$. Then $e_n(z)$ is equal to the number of sequences $(z_0, \cdots, z_n)$ with $z_i - z_{i-1} = -1$ or $+k$

$$(i = 1, \cdots, n), \quad z_0 = 0, \quad z_n = z, \quad z_i \neq z \qquad (i = 0, 1, \cdots, n-1).$$

Further, $H(i)$ is equal to the number of sequences $(z_n, z_{n-1}, \cdots, z_0)$ with

$$n = -\alpha + i(k+1) \geqq 1, \qquad z_i - z_{i-1} = -1 \text{ or } k$$

$(i = 1, \cdots, n)$, $z_n = \alpha$, $z_0 = 0$, $z_i \neq \alpha (i = 0, \cdots, n-1)$. Hence,

$$H_\alpha(i) = e_\alpha(-\alpha + i(k+1))$$

and the above Theorem yields

$$H_\alpha(i) = -\alpha r_\alpha(n)/n + k(k+1)\sum_{0<m\leqq\alpha} r_\alpha(-m)r_{-k}(m+n-1)/(m+n-1),$$

where $n = -\alpha + i(k+1)$. Noting that $r_h(s)$ is equal to the coefficient of w^{k+s} in the expansion of $(1 + w^{k+1})^s$ about 0, formula (1) easily follows.

ABSTRACTS

(Abstracts of papers presented at the Los Angeles Meeting of the Institute, December 27-28, 1957)

1. Non-parametric Multiple-Decision Procedures for Selecting That one of K Populations Which has the Highest Probability of Yielding the Largest Observation. (Preliminary Report) Robert Bechhofer, Cornell University and Milton Sobel, Bell Telephone Laboratories. (By title)

Let X_i be chance variables with density function $f_i(x)$, and let

$$p_i = \text{Prob}\{X_i > \max_{j \neq i} X_j\} \ (i = 1, \cdots, k).$$

Then $\sum_{i=1}^{k} p_i = 1$. Let $p_{[1]} \leq \cdots \leq p_{[k]}$ denote the ranked p_i. Let θ^*, P^* $(1 < \theta^* < \infty,$ $1/k < P^* < 1)$ be specified constants. The goal is to select the population associated with $p_{[k]}$; the procedure must guarantee, (*) Prob {Correct Selection | $p_{[k]} \geq \theta^* p_{[k-1]}\} \geq P^*$. *Procedure*: "At the *m*th stage take the vector-observation $\mathbf{x}_m = (x_{1m}, \cdots, x_{km})$ where the x_{ij} $(j = 1, 2, \cdots)$ are independent observations from the *i*th population. Consider $\mathbf{y}_m = (y_{1m}, \cdots, y_{km})$ which is obtained by replacing the largest component of $\mathbf{x}_m$ by unity, and all other components by zero. Then $\mathbf{y}_m$ is an observation from a multinomial distribution with probability p_i associated with the *i*th component $(i = 1, 2, \cdots, k)$. (*) now can be guaranteed by continuing with procedures already proposed, e.g., these *Annals*, Vol. 27, p. 861. If $f_i(x) = g\{(x - \mu_i)/\theta\}$ $(i = 1, 2, \cdots, k)$, then the procedure can be used for selecting the population associated with the largest μ_i for any θ, known or unknown. Similar non-parametric procedures in which pairs of observations are taken from each population at each stage of experimentation, and which employ the range of each pair can be used for selecting that one of k populations which has the highest probability of yielding the largest sample range. If $f_i(x) = h\{(x - \mu_i)/\theta_i\}$ $(i = 1, 2, \cdots, k)$, then these latter procedures can be used for selecting the population associated with the largest θ_i for any set of μ_i, known or unknown. (Research supported in part by the U. S. Air Force through the Air Force Office of Scientific Research, ARDC, Contract No. AF 18(600)-331.) (Received September 25, 1957.)

2. *The Asymptotic Efficiency of Friedman's χ_r^2-test.* Ph. van Elteren, Mathematical Centre, Amsterdam. (By title)

Let $F(x)$ be a continuous cdf with density function $f(x) = F'(x)$ and let

$$x_{\mu\nu} \ (\mu = 1, 2, \cdots, m, \ \nu = 1, 2, \cdots, n)$$

be a chance variable with distribution $F_{\mu\nu}(x) = F(x + \theta_\nu + \eta_\mu)$. It is assumed for convenience, that $\sum_\nu \theta_\nu = 0$. Friedman (1937) has constructed the χ_r^2-test for the hypothesis $\theta_1 = \theta_2 = \cdots = \theta_n = 0$ (*J. Amer. Stat. Assn.*, Vol. 32, pp. 675-699). For alternatives $\theta_\nu = \theta_{\nu m} = \delta_\nu/\sqrt{m}$, where the δ_ν are given constants satisfying $\sum_\nu \delta_\nu = 0$, the asymptotic relative efficiency for $m \to \infty$ in the sense of Pitman of Friedman's test with respect to the corresponding 2-way-analysis of variance test is found to be $e_n = 12n(n + 1)^{-1}\{\sigma \int f^2(x)\,dx\}^2$, where σ^2 is the variance associated with $F(x)$. If $f(x)$ is normal, e_n reduces to $e_n = 3n/\pi(n + 1)$. (Received August 19, 1957.)

3. Experiments With Mixtures. Henry Scheffé, University of California.

Experiments with mixtures of q components are considered, whose purpose is the empirical prediction of the response to any mixture of the components, when the ...

depends on the proportions $x_1, x_2, \cdots, x_q$ of the components present but not on the total amount. The factor space is then the $(q-1)$-dimensional simplex where $x_1 + \cdots + x_q = 1$, $x_i \geqq 0$. An experimental design called the *simplex lattice* and some modifications are treated; in the simplex lattice $x_i = 0, 1/m, 2/m, \cdots, 1$ for $i = 1, \cdots, q$ and some positive integer m, and the responses of all mixtures possible with these proportions are observed. The usual resolution of the response into general mean, main effects, and interactions does not seem possible, and so polynomial regression is employed. The problem of fitting an nth degree polynomial in $x_1, \cdots, x_q$ to the response is complicated by the fact that different polynomials give the same function on the simplex. Useful canonical forms are developed for $n \leqq 3$. The coefficients in these forms are interpreted as various kinds of synergisms. The analysis of experiments with these designs leads to classes of polynomials orthogonal on the lattices. The paper will appear in *J. Royal Stat. Soc., Series B.* (Received October 25, 1957.)

4. Least-Squares Estimation when Residuals are Correlated. M. M. Siddiqui, University of North Carolina.

Let y_j, $j = 1, \cdots, N$ be observations on a variate and let $y_j = \sum_{i=1}^{p} \beta_i x_{ij} + \Delta_j$, $j = 1, 2, \cdots, N$, where x_{ij} are non-stochastic, and $\Delta' = (\Delta_1, \cdots, \Delta_N)$ is a $N(0, \sigma^2 P)$ vector, where 0 is a zero vector and P is an $N \times N$ correlation matrix. Using the usual least-squares estimates, b_i, of β_i which are obtained by minimizing $\sum \Delta_j^2$, and s^2 of σ^2, the covariance matrix of b_i is obtained for general P and bounds are set on these covariances by first obtaining the maxima and minima of a quadratic and a bilinear form $u'Au$ and $u'Av$ where u and v are $N \times 1$ vectors and A is an $N \times N$ real symmetric matrix under the conditions $u'u = v'v = 1$, $u'v = 0$. (Received October 31, 1957.)

5. A Property of Additively Closed Families of Distributions. Edwin L. Crow, Boulder Laboratories, National Bureau of Standards.

Consider a one-parameter additively closed family of univariate cumulative distribution functions $F(x; \lambda)$ (H. Teicher, *Ann. Math. Stat.*, Vol. 25 (1954), pp. 775–778). Let three cumulants with orders in arithmetic progression exist and be non-zero. If all three orders are even, or if the first order is odd, it is also required that $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$. Consider linear combinations, with real, non-zero coefficients, of a finite number of independent variables with distributions in the family. It is proved that the only such linear combinations whose distributions are also in the family are those with coefficients unity. The additively closed families having this property may be called *strictly additively closed*. It can be shown that (one-parameter) additively closed stable families of distributions (normal and Cauchy in particular) with characteristic functions continuous in λ are not strictly additively closed, while Poisson, generalized Poisson, binomial, and gamma families are strictly additively closed. (Received October 31, 1957.)

6. Determining Sample Size for a Specified Width Confidence Interval. Franklin A. Graybill, Oklahoma State University.

If an experimenter decides to use a confidence interval to locate a parameter, he is concerned with at least two things: (1) Does the interval contain the parameter? (2) How wide is the interval? In general the answer to these questions cannot be given with absolute certainty, but must be given with a probability statement. The problem the experimenter then faces is: The determination of n, the sample size, such that (A) the probability will be equal to α that the confidence interval contains the parameter, and (B) the probability will be equal to β^2 that the width of the confidence interval will be less than d units (where α,

β^2, and d are specified). To solve this problem will generally require two things: (1) The form of the frequency function from which the sample of size n is to be selected; (2) Some previous information on the unknown parameters in the frequency function. This suggests that the sample be taken in two steps; the first sample will be used to determine the number of observations n to be taken in the second sample so that (A) and (B) will be satisfied. For a confidence interval on the mean of a normal population with unknown variance this problem has been solved by Stein for $\beta^2 = 1$. In this paper a theorem is proved which gives a method for determining n so that (A) and (B) will be satisfied. The theorem holds for parameters in the normal distribution and other distributions as well. (Received October 30, 1957.)

7. Nonparametric Estimation of Sample Percentage Point Standard Deviation.

JOHN E. WALSH, Lockheed Aircraft Corporation

The available data consists of a random sample $x(1) < \cdots < x(n)$ from a reasonable well-behaved continuous statistical population. The problem is to estimate the standard deviation of a specified $x(r)$ that is not in the tails of the sample. The estimates examined are of the form $a[x(r + i) - x(r - i)]$ and the explicit problem consists of determining suitable values for a and i. The solution $a = (1/2)(n + 1)^{-3/10}\{[r/(n + 1)][1 - r/(n + 1)]\}^{1/2}$ and $i \doteq (n + 1)^{4/5}$ appears to be satisfactory. Then the expected value of the estimate equals the standard deviation of $x(r)$ plus $O(n^{-9/10})$, also the standard deviation of this estimate is $O(n^{-9/10})$. That is, the fixed and random errors for this point estimate are of the same order of magnitude with respect to n. Solutions can be obtained which decrease the order of one of these types of error. However, these solutions increase the order of the other type of error, so that the over-all error magnitude exceeds $O(n^{-9/10})$. (Received November 7, 1957.)

8. On the Structure of Distribution-Free Statistics.

C. B. BELL, Xavier University of Louisiana and Stanford University

Let $X_1, \cdots, X_n$ be a sample of a one dimensional random variable X which has continuous cpf F. It has been observed that the distribution-free statistics commonly appearing in the literature can be written in the form $\Phi[F(X_1), \cdots, F(X_n)]$, where Φ is a measurable symmetric function defined on the unit cube. Such statistics are said to have structure (d). In establishing that having structure (d) is equivalent to being symmetric and strongly distribution-free for properly closed, symmetrically complete classes of cpf's, this paper extends a result of Birnbaum and Rubin while employing different methodology. These results interest a statistician because (1) they indicate that one should construct a statistic of structure (d) whenever one wishes to design a distribution-free statistic; and (2) they guarantee that each symmetric, strongly distribution-free statistic is of structure (d), and, hence, that the value of its cpf at any point is the volume of a polyhedral region in the unit cube. Under such circumstances the work of numerous statisticians indicate that it should be possible to evaluate the cpf explicitly, reduce it to recursion formulae; tabulate it with high-speed computers, or evaluate its limiting distribution. (Received November 7, 1957.)

9. On the Supremum of the Poisson Process.

RONALD PYKE, Stanford University.

Let $\{X(t); t \geqq 0\}$ be a Poisson process (with shift) for which $\log E(e^{iwX(t)}) = -itw\alpha + \lambda t(e^{iw} - 1)$, $w \in R_1$, $\alpha, \lambda > 0$. Define $\sigma(x, T) = Pr\left\{\sup_{0 < t \leq T} X(t) \leqq x\right\}$. Let $X_1, X_2, \cdots X_n$ be the ordered random variables of n independent and uniform $-$ (0, 1) random

variables. The distribution function, $Pr\left\{\max_{1 \leqq i \leqq n} (ai - X_i) \leqq x\right\}$, is obtained for all a, $x \in R_1$. For $a = 1/n$, this reduces to the distribution function of D_n^+ (cf. Birnbaum and Tingey, A.M.S., Vol. 22). Utilizing this result, $\sigma(x, T)$ is obtained explicitly. Applications of these expressions to queueing theory and distribution-free statistics are given.

10. On the Distributions of Various Sums of Squares in an Analysis of Variance Table for Different Classifications With Correlated and Non-homogeneous Errors. B. R. Bhat, Karnatak University. (Preliminary Report) (By Title)

The distributions of various sums of squares in an analysis of variance table for two way classification have been obtained by Box (Ann. Math. Stat., Vol. 25, pp. 484–498) under the assumption that the vectors $X_{.j}$ for $j = 1, 2, \cdots q$ are independent vector observations from a p-variate normal population with mean μ and covariance matrix Σ. The vector $X_{.j}$ for each jth level of the factor B denotes p observations corresponding to the p levels of the other factor A. This paper gives the distributions of the various sums of squares for any n-way classification under similar normality assumptions. It is noted that these distributions, in general, follow a simple pattern and so is their mutual dependence. For $n = 3$, if we have a third factor C at r levels in addition to the above factors A and B and if we assume that $X_{..k}$ for $k = 1, 2, \cdots r$ are independent vector observations from a pq-variate normal population, then, according to the general pattern the first set consists of distributions of the sums of squares for the correction term, main effects A and B and their interaction $A\,B$. The second set (the only remaining set) consists of the distributions of the sums of squares for the remaining main factor C and its interactions with the effects in the first set. Any two distributions, not belonging to the same set are independent, whereas, the distributions in the same set are mutually dependent. (Received May 10, 1957)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of The Institute news items of interest

Personal Items

During 1957–58 T. W. Anderson will be a Fellow at the Center for Advanced Study in the Behavioral Sciences in Stanford, California.

Dr. Robert S. Aries is now Chairman of the Board of Aries Associates, whose offices for general consultation were recently transferred to 77 South Street, Stamford, Connecticut.

John Bailey has recently joined the staff of the Waltham Laboratories of Sylvania Electric Products, Inc., as an Engineer in their Applied Engineering Department.

Colin R. Blyth, on leave from the University of Illinois, will be at Stanford University for the academic year of 1957–58.

John V. Breakwell has taken a position as Staff Scientist with the Lockheed Missile Systems Division in Palo Alto.

D. M. Brown is now studying for the degree of Ph.D. in statistics at Princeton University on a RAND Corporation Fellowship.

Charles R. Carr has joined the research staff of The RAND Corporation at Santa Monica, California.

Richard L. Carter has received his Ph.D degree in statistics from the University of North Carolina and has been appointed Associate Professor of Industrial Engineering at the Illinois Institute of Technology.

Jonas M. Dalton completed his work for his Master's degree in statistics at Virginia Polytechnic Institute in June, 1957. He is now employed at the Bell Telephone Laboratories, Murray Hill, New Jersey.

Morris H. De Groot has been appointed Assistant Professor in the Department of Mathematics at Carnegie Institute of Technology

R. F. Drenick has joined the Bell Telephone Laboratories as a member of its technical staff.

Joseph Dubay is now an instructor in the Mathematics Department of the University of Oregon.

Professor Benjamin Epstein of Wayne State University is on leave at the Department of Statistics, Stanford University

L. A. Gardner, Jr., has resigned his position as research scientist at Columbia University's Hudson Laboratories and is now employed as staff mathematician at M.I.T. Lincoln Laboratory

John J. Gart is now a graduate fellow at the Oak Ridge Institute of Nuclear Studies, continuing work there toward a Ph D in statistics from V.P.I.

David W. Gaylor has resigned from the Nuclear Aircraft Research Facility, Convair, Ft. Worth, to work toward a Ph D. in experimental statistics at North Carolina State College.

R. Gnanadesikan is now working with the Statistics Group at the Proctor and Gamble Company at Cincinnati, Ohio

William A. Golomski, formerly Assistant Professor of Mathematics at Marquette University, is now in charge of operations research for Oscar Mayer and Company, Inc., Madison, Wisconsin.

Roe Goodman has gone from Santiago, Chile, where he was F.A.O. agricultural statistician, to Karachi, Pakistan, where he is now sampling statistician under the I.C.A. program of the U S. Government in that country.

Ulf Grenander has accepted an appointment as Professor of Mathematical Statistics at Brown University.

Irwin Guttman is on leave of absence from the University of Alberta and will spend the academic year of 1957–58 as a Research Associate in the Department of Mathematics, Statistical Section, of Princeton University.

Following completion of assignment for Remington Rand Internation (installation of UNIVAC I at European Computing Center, Frankfurt/Main. Germany), Dr. Carl Hammer has accepted a similar position with Sylvania Electric Products, Inc., at their Waltham Laboratories.

Gordon M. Harrington has left his position as consultant in research, Connecticut State Department of Education, to become Associate Professor of Psychology and Department Chairman at Wilmington College, Wilmington, Ohio.

variables. The distribution function, $Pr\left\{\max_{1 \leqq i \leqq n} (ai - X_i) \leqq x\right\}$, is obtained for all $a, x \in R_1$. For $a = 1/n$, this reduces to the distribution function of D_n^+ (cf. Birnbaum and Tingey, A.M.S., Vol. 22). Utilizing this result, $\sigma(x, T)$ is obtained explicitly. Applications of these expressions to queueing theory and distribution-free statistics are given.

10. On the Distributions of Various Sums of Squares in an Analysis of Variance Table for Different Classifications With Correlated and Non-homogeneous Errors. B. R. Bhat, Karnatak University. (Preliminary Report) (By Title)

The distributions of various sums of squares in an analysis of variance table for two way classification have been obtained by Box (Ann. Math. Stat., Vol. 25, pp. 484–498) under the assumption that the vectors $X_{.j}$ for $j = 1, 2, \cdots q$ are independent vector observations from a p-variate normal population with mean μ and covariance matrix Σ. The vector $X_{.j}$ for each jth level of the factor B denotes p observations corresponding to the p levels of the other factor A. This paper gives the distributions of the various sums of squares for any n-way classification under similar normality assumptions. It is noted that these distributions, in general, follow a simple pattern and so is their mutual dependence. For $n = 3$, if we have a third factor C at r levels in addition to the above factors A and B and if we assume that $X_{..k}$ for $k = 1, 2, \cdots r$ are independent vector observations from a pq-variate normal population, then, according to the general pattern the first set consists of distributions of the sums of squares for the correction term, main effects A and B and their interaction $A B$. The second set (the only remaining set) consists of the distributions of the sums of squares for the remaining main factor C and its interactions with the effects in the first set. Any two distributions, not belonging to the same set are independent, whereas, the distributions in the same set are mutually dependent. (Received May 10, 1957)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of The Institute news items of interest

Personal Items

During 1957–58 T. W. Anderson will be a Fellow at the Center for Advanced Study in the Behavioral Sciences in Stanford, California.

Dr. Robert S. Aries is now Chairman of the Board of Aries Associates, whose offices for general consultation were recently transferred to 77 South Street, Stamford, Connecticut.

John Bailey has recently joined the staff of the Waltham Laboratories of Sylvania Electric Products, Inc., as an Engineer in their Applied Engineering Department.

Colin R. Blyth, on leave from the University of Illinois, will be at Stanford University for the academic year of 1957–58.

John V. Breakwell has taken a position as Staff Scientist with the Lockheed Missile Systems Division in Palo Alto.

D. M. Brown is now studying for the degree of Ph.D. in statistics at Princeton University on a RAND Corporation Fellowship.

Charles R. Carr has joined the research staff of The RAND Corporation at Santa Monica, California.

Richard L. Carter has received his Ph.D. degree in statistics from the University of North Carolina and has been appointed Associate Professor of Industrial Engineering at the Illinois Institute of Technology.

Jonas M. Dalton completed his work for his Master's degree in statistics at Virginia Polytechnic Institute in June, 1957 He is now employed at the Bell Telephone Laboratories, Murray Hill, New Jersey.

Morris H. De Groot has been appointed Assistant Professor in the Department of Mathematics at Carnegie Institute of Technology

R. F. Drenick has joined the Bell Telephone Laboratories as a member of its technical staff.

Joseph Dubay is now an instructor in the Mathematics Department of the University of Oregon.

Professor Benjamin Epstein of Wayne State University is on leave at the Department of Statistics, Stanford University.

L. A. Gardner, Jr., has resigned his position as research scientist at Columbia University's Hudson Laboratories and is now employed as staff mathematician at M.I.T. Lincoln Laboratory.

John J. Gart is now a graduate fellow at the Oak Ridge Institute of Nuclear Studies, continuing work there toward a Ph D in statistics from V.P.I.

David W. Gaylor has resigned from the Nuclear Aircraft Research Facility, Convair, Ft. Worth, to work toward a Ph D in experimental statistics at North Carolina State College.

R. Gnanadesikan is now working with the Statistics Group at the Proctor and Gamble Company at Cincinnati, Ohio.

William A. Golomski, formerly Assistant Professor of Mathematics at Marquette University, is now in charge of operations research for Oscar Mayer and Company, Inc., Madison, Wisconsin.

Roe Goodman has gone from Santiago, Chile, where he was F.A.O. agricultural statistician, to Karachi, Pakistan, where he is now sampling statistician under the I.C.A. program of the U. S. Government in that country.

Ulf Grenander has accepted an appointment as Professor of Mathematical Statistics at Brown University.

Irwin Guttman is on leave of absence from the University of Alberta and will spend the academic year of 1957–58 as a Research Associate in the Department of Mathematics, Statistical Section, of Princeton University.

Following completion of assignment for Remington Rand Internati[illegible] stallation of UNIVAC I at European Computing Center, Frankfurt [illegible] Germany), Dr. Carl Hammer has accepted a similar position with [illegible] Electric Products, Inc., at their Waltham Laboratories.

Gordon M. Harrington has left his position as consultant in [illegible] icut State Department of Education, to become Associate [illegible] chology and Department Chairman at Wilmington College [illegible]

Dr. Theodore W. Horner, formerly of the Statistical Laboratory, Iowa State College, is now an Operations Research Analyst with General Mills, Inc., Minneapolis, Minnesota.

Patricia A. Inman received her M. A. in mathematics at U.C.L.A. in August, 1957, and has accepted a position as Computing Engineer at Atomics International, Canoga Park, California.

James Edward Jackson has taken a leave of absence from the Eastman Kodak Company to work on a doctorate at V.P.I. in Blacksburg, Virginia.

Trinidad J. Jaramillo is now working with System Research, University of Chicago, as Senior Research Engineer.

Peter W. M. John has accepted a position as research statistician with the California Research Corporation at Richmond, California.

Andre G. Laurent, formerly with the Department of Statistics of Michigan State University, has accepted an appointment as Associate Professor in the Department of Mathematics of Wayne State University.

J. Walter Lynch has moved from Huntsville, Alabama, to 101-L Rodman Road, Aberdeen, Maryland.

Albert Madansky has been employed by the Mathematics Division, The RAND Corporation, Santa Monica, California.

B. Mandelbrot, formerly of the University of Geneva, has accepted an appointment in the University of Lille.

Mr. C. L. Marcus has returned to the University of Illinois to complete his studies for a Ph.D. in statistics. He also holds an Assistantship in the Mathematics Department and a Fellowship from Armour Research.

Dr. H. P. Mulholland has been appointed to a Senior Lectureship in Mathematics in the University of Exeter.

Peter Newman has taken a post as Lecturer in Economics University College of the West Indies, Mona, St. Andrew, Jamaica, B.W.I.

Dr. Bernard Ostle, formerly Professor of Mathematics and Director of the Statistical Laboratory at Montana State College, is now with the Reliability Department of Sandia Corporation, Albuquerque, New Mexico.

Dr. Raymond P. Peterson has accepted a position as Mathematical Statistician with the Research Department, Matson Navigation Company, San Francisco, California.

Paul H. Randolph has resigned his position as Assistant Professor of Industrial Engineering at Illinois Institute of Technology to accept a position as Associate Professor of Industrial Engineering at Purdue University.

Dr. George J. Resnikoff, formerly Research Associate at the Applied Mathematics and Statistics Laboratory, Stanford University, has joined the Industrial Engineering Department of the Illinois Institute of Technology as Associate Professor.

Robert H. Riffenburgh received his Ph.D. degree in statistics at the Virginia Polytechnic Institute and is now Assistant Professor of Mathematics at the University of Hawaii, Honolulu.

Joseph Rosenbaum has resigned his position as Associate Mathematician, Systems Development Division, The RAND Corporation, and is presently employed as statistician, Broadview Research Corporation, Burlingame, California.

Dr. Jagdish S. Rustagi is an Assistant Professor in the Department of Statistics at Michigan State University, East Lansing, Michigan, for the current academic year.

Melvin E. Salveson has formed the Council for Advanced Management together with Herbert Holt, M.D., and is offering services for research in management and management education.

Dr. M. M. Siddiqui, who obtained his doctorate degree from the University of North Carolina in June, 1957, is now working for a temporary period with the Boulder Laboratories of the National Bureau of Standards in the capacity of Mathematical Statistician.

Professor Jack Wilber has returned to Roosevelt University after spending four months as Consultant to the Operations Analysis Office at the Air Force Missile Test Center.

Morris Skibinsky has returned to the Statistical Laboratory at Purdue after a year's leave of absence at Michigan State University.

James H. Stapleton received his Ph.D. degree in mathematical statistics from Purdue University in June, 1957, and is now a statistical consultant in the statistical methods section of General Electric's General Engineering Laboratory in Schenectady, New York.

Daniel Teichroew has joined the Graduate School of Business at Stanford University as Associate Professor of Management.

James E. Thompson has returned to his job as mathematician with the Defense Department, having completed a year of graduate study with the Department of Statistics at Stanford University on a Defense Department fellowship.

W. A. Thompson, Jr., has left the U. S. Air Defense Board and has accepted an academic position at the University of Delaware.

Dr. Fred H. Tingey, mathematician, has been appointed by TECHNICAL OPERATIONS, INCORPORATED, as Assistant Chairman of Experiment Planning and Execution, Dr. Ian W. Tervet, Director of the research and development firm's West Coast office, announced today. Dr. Tingey received his masters and doctoral degrees in Mathematics and Mathematical Statistics at the University of Washington. He graduated from Utah State College in 1947. In his new position, Dr. Tingey will plan and direct field experiments conducted by TECHNICAL OPERATIONS in conjunction with the Combat Development Experimentation Center (CDEC) at Fort Ord, California.

John A. Tischendorf, having completed two years of active duty with the Commissioned Corps, U. S. Public Health Service, has joined the staff of the Allentown Laboratory of Bell Telephone Laboratories, Inc.

Joseph B. Tysver, formerly an Associate Research Engineer at the University of Michigan, received his Ph. D. degree from that University in June, 1957.

and has accepted a position as Research Specialist in the Pilotless Aircraft Division of Boeing Airplane Company, Seattle, Washington.

Dr. John S. White has accepted a position as statistician with the Aero Division of Minneapolis-Honeywell Regulator Company.

Robert A. Wijsman is now Assistant Professor in the statistics group of the Department of Mathematics at the University of Illinois.

New Members

The following persons have been elected to membership in The Institute

August 7, 1957, to November 1, 1957

Abraham, T. C., M. Sc. (Karnatak Univ., India), Teaching Fellow, Boston University Graduate School, Boston University, Boston 15, Massachusetts; *627 Commonwealth Ave., Boston 15, Mass.*

Albert, Arthur E., M. S. (Stanford Univ.), Student, Department of Statistics, Stanford University, Stanford, California; *20 Russell Ave., Portola Valley, Calif.*

Ali, Asghar, M. A. (Univ. of North Carolina), Lecturer in Statistics, *Institute of Statistics, University of the Panjal, Lahore, Pakistan.*

Anglin, Ernie LaRue, B. S. (Univ. of Georgia), Student, *Department of Mathematics, University of Georgia, Athens, Georgia.*

Beatty, Richard L., M. S. (Univ. of Colorado), *Instructor in Statistics, University of Wyoming, Laramie, Wyoming.*

Bland, Richard P., B. S. (Univ. of Oklahoma), Student, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina; *1 Audley Lane, Chapel Hill, N. C.*

Bobotek, Henry G., M. A. (Univ. of Illinois), Research Associate, *Control Systems Laboratory, University of Illinois, Urbana, Illinois.*

Bogdanoff, J. L., Ph. D. (Columbia Univ.), *Professor of Engineering Sciences, Purdue University, Lafayette, Indiana.*

Calhoun, David W., B. A. (Yale Univ.), Biometrician, G. D. Searle and Co., P. O. Box 5110, Chicago 80, Illinois; *820 Hamlin St., Evanston, Ill.*

Caspers, James W., M. S. in E. E. (Univ. of Washington), Head, Applied Theoretical Studies Group, U. S. Navy Electronics Lab., San Diego 52, California; *5014 August St., San Diego 10, Calif.*

Champernowne, D. G., M. A. (Cambridge Univ.), Professor in Statistics and Fellow of Nuffield College, *Oxford University, Nuffield College, Oxford, England.*

Chapman, James W., M. S. (Univ. of Minnesota), Research Assistant, *Department of Soils, Institute of Agriculture, University of Minnesota, St. Paul 1, Minnesota.*

Clarke, Geoffrey M., M. A. (Oxon), Statistician, Department of Agriculture and Horticulture, National Fruit and Cider Institute, University of Bristol; *University of Bristol, Research Station, Long Ashton, Bristol, England.*

Cotton, John W., Ph. D. (Indiana Univ.), Assistant Professor of Psychology, Northwestern University, Evanston, Illinois; *Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois.*

Cox, Constance E., M. S. (Iowa State College), *Head, Biometrics Section, Food and Drug Directorate, Department of National Health and Welfare, Tunney's Pasture, Ottawa, Ontario, Canada.*

Dear, Robert E., Ph. D. (Univ. of Washington), Research Associate, *Research Division, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.*

Engler, Jean, Ph. D. (Northwestern Univ.), National Science Foundation Postdoctoral

Fellow, *Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.*

Evans, Richard V., A. B. (Princeton Univ.), Student, Department of Industrial Engineering, Johns Hopkins University, Baltimore 18, Md., *Barroll Road, Baltimore 9, Md.*

Feldt, Leonard S., Ph. D. (State Univ. of Iowa), Assistant Professor, *College of Education, State University of Iowa, Iowa City, Iowa.*

Fields, Raymond I., M. A. (Univ. of Arizona), Student, Virginia Polytechnic Institute, Blacksburg, Virginia; *Speed Scientific School, University of Louisville, Louisville, Kentucky.*

Figge, Harry J., President, *Harry J. Figge and Associates, 800 Liberty Building, Des Moines 9, Iowa*

Flanagan, Richard M., B. A. (Univ of Michigan), Senior Programmer, *Argonaut Insurance Group, 250 Middlefield Road, Menlo Park, California.*

Gordon, M. H., Ph. D. (Univ. of Tennessee), Assistant Director, Central Neuropsychiatric Research Unit, Veterans Administration Hospital, Perry Point, Maryland; *Box 546, Perry Point, Maryland.*

Govindarajulu, Z., M. A. (Univ. of Madras), Graduate Student, Statistics, and Teaching Assistant, *Biostatistics Division, University of Minnesota, Minneapolis 14, Minnesota*

Green, Lloyd G., B. A. (Washington Missionary College), Mathematician, *Touche, Niven, Bailey and Smart, 1380 National Bank Building, Detroit 26, Michigan.*

Halton, John H., M. A. (Oxon), Research-Student in Faculty of Physical Sciences, *Balliol College, Oxford University, Oxford, England*

Hancock, John V., B. S. (Memphis State Univ), Research Assistant, *Department of Mathematics, University of Georgia, Athens, Georgia*

Harrison, Gerald, Ph. D. (Calif. Institute of Technology), Mathematician, *The Teleregister Corp., 445 Fairfield Avenue, Stamford, Connecticut*

Heinhold, Josef, Dr. rer. nat. (Technische Hochschule München), Professor, *Institut fur Angewandte Mathematik, Technische Hochschule Munchen München 2, NW, Arcisstraße 21, Germany.*

Hicks, Charles R., Ph D. (Syracuse Univ.), Associate Professor of Mathematics and Research Associate in the Statistical Laboratory, *Statistical Laboratory, Engineering Administration Building, Purdue University, Lafayette, Indiana*

Hoyland, Arnljot, Cand. real (Univ. of Oslo), Research Assistant, Forsikringsteknish Seminar, University of Oslo, Blindern pr Oslo, Norway, *Krokrolden 20, Stabekk pr. Oslo, Norway*

Iversen, Iver Andrew, B. A. (Univ. of Minnesota), Teaching Assistant, University of Minnesota, Minneapolis 14, Minnesota; *420 Fifth Street S E , Minneapolis 14, Minnesota*

Jacobsen, Fred M., Jr., Ph D (Iowa State College), Group Leader, Computer Programming and Mathematical Analysis, American Oil Co , Box 401, Texas City, Texas; *Box 1537, Texas City, Texas.*

Jones, Alfred Welwood, Ph. D (Columbia Univ), Systems Engineer, *Bell Telephone Laboratories, 463 West Street, New York 14, N. Y*

Kakeshita, Shin'ichi, B. Sc (Kyushu Univ), Student, *Math Inst . Fac. Sci , Kyushu University, Fukuoka, Japan.*

Kim, Dong Sie, B. S. (Seoul National Univ), Assistant, Dept of Mathematics, Seoul National University, Seoul, Korea; *11-44 Ka hoe-Dong, Chong no Ku, Seoul, Korea.*

Knapp, Leslie E., B. S. and B. A. (Upper Iowa Univ), Student, Stanford University, Stanford, California; *1255 Tucson Avenue, Sunnyvale, California*

Lamm, Richard A., M. A (Hofstra College), Analytical Statistician, Chemical Corps R and D Command, Biological Warfare Laboratories, Fort Detrick, Frederick, Maryland, *75 Stewart Manor, Frederick, Maryland.*

Laubscher, N. F., M. Sc. (Potchefstroomse Universiteit vir C H O), Research Officer, *South African Council for Scientific and Industrial Research, National Physical Research Laboratory, P. O. Box 395, Pretoria, South Africa*

and has accepted a position as Research Specialist in the Pilotless Aircraft Division of Boeing Airplane Company, Seattle, Washington.

Dr. John S. White has accepted a position as statistician with the Aero Division of Minneapolis-Honeywell Regulator Company.

Robert A. Wijsman is now Assistant Professor in the statistics group of the Department of Mathematics at the University of Illinois.

New Members

The following persons have been elected to membership in The Institute

August 7, 1957, to November 1, 1957

Abraham, T. C., M. Sc. (Karnatak Univ., India), Teaching Fellow, Boston University Graduate School, Boston University, Boston 15, Massachusetts; *627 Commonwealth Ave., Boston 15, Mass.*

Albert, Arthur E., M. S. (Stanford Univ.), Student, Department of Statistics, Stanford University, Stanford, California; *20 Russell Ave., Portola Valley, Calif.*

Ali, Asghar, M. A. (Univ. of North Carolina), Lecturer in Statistics, *Institute of Statistics, University of the Panjal, Lahore, Pakistan.*

Anglin, Ernie LaRue, B. S. (Univ. of Georgia), Student, *Department of Mathematics, University of Georgia, Athens, Georgia.*

Beatty, Richard L., M. S. (Univ. of Colorado), *Instructor in Statistics, University of Wyoming, Laramie, Wyoming.*

Bland, Richard P., B. S. (Univ. of Oklahoma), Student, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina; *1 Audley Lane, Chapel Hill, N. C.*

Bobotek, Henry G., M. A. (Univ. of Illinois), Research Associate, *Control Systems Laboratory, University of Illinois, Urbana, Illinois.*

Bogdanoff, J. L., Ph. D. (Columbia Univ.), *Professor of Engineering Sciences, Purdue University, Lafayette, Indiana.*

Calhoun, David W., B. A. (Yale Univ.), Biometrician, G. D. Searle and Co., P. O. Box 5110, Chicago 80, Illinois; *820 Hamlin St., Evanston, Ill.*

Caspers, James W., M. S. in E. E. (Univ. of Washington), Head, Applied Theoretical Studies Group, U. S. Navy Electronics Lab., San Diego 52, California; *5014 August St., San Diego 10, Calif.*

Champernowne, D. G., M. A. (Cambridge Univ.), Professor in Statistics and Fellow of Nuffield College, *Oxford University, Nuffield College, Oxford, England.*

Chapman, James W., M. S. (Univ. of Minnesota), Research Assistant, *Department of Soils, Institute of Agriculture, University of Minnesota, St. Paul 1, Minnesota.*

Clarke, Geoffrey M., M. A. (Oxon), Statistician, Department of Agriculture and Horticulture, National Fruit and Cider Institute, University of Bristol; *University of Bristol, Research Station, Long Ashton, Bristol, England.*

Cotton, John W., Ph. D. (Indiana Univ.), Assistant Professor of Psychology, Northwestern University, Evanston, Illinois; *Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois.*

Cox, Constance E., M. S. (Iowa State College), *Head, Biometrics Section, Food and Drug Directorate, Department of National Health and Welfare, Tunney's Pasture, Ottawa, Ontario, Canada.*

Dear, Robert E., Ph. D. (Univ. of Washington), Research Associate, *Research Division, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.*

Engler, Jean, Ph. D. (Northwestern Univ.), National Science Foundation Postdoctoral

Fellow, *Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.*

Evans, Richard V., A. B. (Princeton Univ), Student, Department of Industrial Engineering, Johns Hopkins University, Baltimore 18, Md , *Barroll Road, Baltimore 9, Md.*

Feldt, Leonard S., Ph. D. (State Univ. of Iowa), Assistant Professor, *College of Education, State University of Iowa, Iowa City, Iowa.*

Fields, Raymond I., M. A. (Univ. of Arizona), Student, Virginia Polytechnic Institute, Blacksburg, Virginia; *Speed Scientific School, University of Louisville, Louisville, Kentucky.*

Figge, Harry J., President, *Harry J. Figge and Associates, 800 Liberty Building, Des Moines 9, Iowa*

Flanagan, Richard M., B. A. (Univ. of Michigan), Senior Programmer, *Argonaut Insurance Group, 250 Middlefield Road, Menlo Park, California*

Gordon, M. H., Ph. D. (Univ. of Tennessee), Assistant Director, Central Neuropsychiatric Research Unit, Veterans Administration Hospital, Perry Point, Maryland; *Box 546, Perry Point, Maryland.*

Govindarajulu, Z., M. A. (Univ of Madras), Graduate Student, Statistics, and Teaching Assistant, *Biostatistics Division, University of Minnesota, Minneapolis 14, Minnesota*

Green, Lloyd G., B. A. (Washington Missionary College), Mathematician, *Touche, Niven, Bailey and Smart, 1380 National Bank Building, Detroit 26, Michigan.*

Halton, John H., M. A. (Oxon), Research-Student in Faculty of Physical Sciences, *Balliol College, Oxford University, Oxford, England*

Hancock, John V., B. S. (Memphis State Univ), Research Assistant, *Department of Mathematics, University of Georgia, Athens, Georgia*

Harrison, Gerald, Ph. D. (Calif Institute of Technology), Mathematician, *The Teleregister Corp., 445 Fairfield Avenue, Stamford, Connecticut*

Heinhold, Josef, Dr. rer. nat. (Technische Hochschule München), Professor, *Institut fur Angewandte Mathematik, Technische Hochschule Munchen Munchen 2, NW, Arcisstraße 21, Germany.*

Hicks, Charles R., Ph. D. (Syracuse Univ), Associate Professor of Mathematics and Research Associate in the Statistical Laboratory, *Statistical Laboratory, Engineering Administration Building, Purdue University, Lafayette, Indiana.*

Hoyland, Arnljot, Cand. real (Univ of Oslo), Research Assistant, Forsikringsteknisk Seminar, University of Oslo, Blindern pr Oslo, Norway, *Krokolden 20, Stabekk pr. Oslo, Norway*

Iversen, Iver Andrew, B. A. (Univ of Minnesota), Teaching Assistant, University of Minnesota, Minneapolis 14, Minnesota, *420 Fifth Street S E., Minneapolis 14, Minnesota.*

Jacobsen, Fred M., Jr., Ph. D. (Iowa State College), Group Leader, Computer Programming and Mathematical Analysis, American Oil Co , Box 401, Texas City, Texas; *Box 1537, Texas City, Texas.*

Jones, Alfred Welwood, Ph. D. (Columbia Univ), Systems Engineer, *Bell Telephone Laboratories, 463 West Street, New York 14, N. Y*

Kakeshita, Shin'ichi, B. Sc. (Kyushu Univ.), Student, *Math. Inst , Fac. Sci , Kyushu University, Fukuoka, Japan.*

Kim, Dong Sie, B. S. (Seoul National Univ), Assistant, Dept. of Mathematics, Seoul National University, Seoul, Korea; *11-44 Ka-heo-Dong, Chong-no-Ku, Seoul, Korea.*

Knapp, Leslie E., B. S. and B. A. (Upper Iowa Univ.), Student, Stanford University, Stanford, California; *1255 Tucson Avenue, Sunnyvale, California.*

Lamm, Richard A., M. A. (Hofstra College), Analytical Statistician, Chemical Corps R and D Command, Biological Warfare Laboratories, Fort Detrick, Frederick, Maryland, *75 Stewart Manor, Frederick, Maryland.*

Laubscher, N. F., M. Sc. (Potchefstroomse Universiteit vir C.H O.), A. Research Officer, *South African Council for Scientific and Industrial Research, National Physical Research Laboratory, P. O. Box 395, Pretoria, South Africa.*

Lindeman, Richard H., M. S. (Univ. of Wisconsin), Research Fellow, Bureau of Institutional Research, University of Minnesota, Minneapolis 14, Minnesota; *4936 Penn Ave. South, Minneapolis 9, Minn.*

Lokki, Olli Kristian, Dr. phil. (Univ. of Helsinki), *Associated Professor, Institute of Technology, Helsinki, Finland.*

Lunneborg, Clifford E., B. S. (Univ. of Washington), Research Assistant, Division of Counseling and Testing Services, University of Washington, Seattle, Washington; *5218 16th Avenue N.E., Seattle 5, Washington.*

McGuire, Judson Ulery, Jr., Ph. D. (Iowa State College) Entomologist, Agricultural Research Service, U. S. Department of Agriculture, Washington, D. C.; Apartado 654, Camaguey, Cuba.

Mikami, Misao, D. Sc. (Kyushu Univ.), Professor of Industrial Statistics, *Seminar of Industrial Statistics, Faculty of Engineering, Kyushu University, Fukuoka, Japan.*

Mitra, S. S., M. S. (Univ. of Calcutta), Graduate Student and Teaching Assistant, *Department of Mathematics, University of Washington, Seattle 5, Washington.*

Miyasawa, Koichi, D. Sc. (Kyushu Univ.), Assistant Professor of Mathematical Statistics and Econometrics, *Faculty of Economics, Tokyo University, Tokyo, Japan.*

Pendergrass, R. N., M. A. (Univ. of Missouri), *Professor of Mathematics, Radford College, Radford, Virginia.*

Pike, M. C., B. S. (Witwatersrand Univ., South Africa), Student, Statistical Laboratory, University of Cambridge, *Trinity Hall, Cambridge, England.*

Redus, Faye, B. S. (Stephen F. Austin State College), Senior Analyst and Programmer, Sutherland Co., Suite 1112, First National Bank Building, Peoria, Illinois; *434 W. 21 Street, San Bernardino, California.*

Reed, James C., Ph. D. (Univ. of Chicago), *Director, Reading and Study Skills, Wayne State University, Detroit, Michigan.*

Rice, Philip L., B. S. (Principia College), Chief, Tropospheric Analysis Section, Radio Propagation Engineering Division, National Bureau of Standards, Boulder, Colorado; *1103 Pine Street, Boulder, Colorado.*

Rogerson, G. W., B. S. (Melbourne), Student, Melbourne University, Carlton, Melbourne N3, Victoria, Australia; *48 Drummond St., Carlton, Melbourne N3, Victoria, Australia.*

Romano, Albert, M. A. (Washington Univ.), Student and Research Assistant, *Dept. of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia.*

Sagi, Philip C., Ph. D. (Univ. of Minnesota), Research Associate, *Office of Population Research, 5 Ivy Lane, Princeton, New Jersey.*

Schoderbek, Joseph J., M. S. (Carnegie Inst. of Tech.), Research Engineer, Missile Systems Division, Lockheed Aircraft Corp., Palo Alto, California; *543 Carla Court, Mountain View, Calif.*

Schwartz, A. J., B. S. (Wayne Univ.), Student, Wayne State University, Detroit, Michigan; *18478 Prest, Detroit, Michigan.*

Shaffer, Douglas H., Ph. D. (Carnegie Inst. of Technology), Mathematician, *Westinghouse Research, Pittsburgh 35, Pa.*

Sherman, Seymour, Ph. D. (Cornell Univ.), Professor, *Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia 4, Pennsylvania.*

Singh, Rajinder, M. A. (Panjale Univ., India), Graduate Assistant, *University of Illinois, Department of Mathematics, Urbana, Illinois.*

Spicer, Ira G., B. S. (Univ. of Minnesota), Development Engineer, Project Leader of Technical Analysis, Minneapolis-Honeywell, Ordnance Engineering, Hopkins, Minnesota; 1928 Emerson Ave. So., Apt. 1-D, *Minneapolis 5, Minnesota.*

Trammell, Carol D., B. S. (Carnegie Inst. of Tech.), Graduate Student and Teaching Assistant, *Department of Mathematics, Carnegie Institute of Technology, Pittsburgh 13, Pennsylvania.*

Weiler, H., M. S. (N.S.W. Univ. of Technology), Research Officer, *CSIRO, Sheep Biology Laboratory, P. O. Box 144, Paramatta, N.S.W., Australia.*

Woods, W. Max, M. S. (Oregon Univ.), Student, Stanford University, Stanford, California; *1919 Manhattan Ave., Apt. 4, East Palo Alto, California.*
Youtcheff, John S., A. B. (Columbia Univ), Operations Analyst, General Electric Company, Missile and Ordnance Systems Dept , Philadelphia, Pennsylvania; *Post Office Box 165, Berwyn, Pennsylvania.*

Congratulations to the Office of Naval Research

At the request of the Council, the President of the Institute of Mathematical Statistics has written a congratulatory letter to Admiral Bennett of the Office of Naval Research in connection with the tenth anniversary of the Office of Naval Research. The text of the letter follows

Dear Admiral Bennett:

At the recent annual meeting of the Institute of Mathematical Statistics, the Council unanimously asked me to offer our congratulations and best wishes on the tenth anniversary of the establishment of the Office of Naval Research.

Through its support of senior investigators and graduate students and the consequent publication of many important technical papers and books, the Office of Naval Research has been contributing greatly to the advancement of fundamental research in mathematical statistics and probability theory. This contribution is especially important because it is being made during a period when these fields are showing themselves capable of particularly rapid growth.

The help you have given our profession is but one aspect of that program through which government and science work hand in hand to the benefit of each and to that of the nation as a whole, both in military and civilian pursuits. The Navy Department is outstanding in this respect, and it must be pleased and honored by the record of your Office.

Congratulations to you and your staff and best wishes for many more years of success in your efforts.

Respectfully yours,
Leonard J. Savage
President

Fifth Annual Southern Regional Graduate Summer Session in Statistics

The fifth Southern Regional Graduate Summer Session in Statistics will be held June 16 through July 26, 1958, at Oklahoma State University, Stillwater, Oklahoma. The summer sessions are rotated annually among the following institutions: Virginia Polytechnic Institute, Oklahoma State University, University of Florida and North Carolina State College.

The program may be entered at any session, and consecutive courses will be offered in successive summers. The summer work in statistics may be appl

towards residence requirements at any one of the cooperating institutions, as well as certain other institutions, in partial fulfillment of residence requirements for graduate degrees. Each annual summer session lasts six weeks and the several courses offered carry three semester hours of graduate credit.

The summer sessions are designed to carry out a recommendation of the Southern Regional Education Board's Committee on Statistics, on which the four institutions initiating the program are represented.

The sessions will be of particular interest to (1) research and professional workers who want intensive instruction in basic statistical concepts and who wish to learn modern statistical methodology, (2) teachers of elementary statistics courses who want some formal training in modern statistics, (3) prospective candidates for graduate degrees in statistics, (4) graduate students in other fields who desire supporting work in statistics, and (5) professional statisticians who wish to keep informed of advanced specialized theory and methods.

The faculty for the 1958 Summer Session at Oklahoma State University will include the following visiting professors: H. O. Hartley, Statistical Laboratory, Iowa State College; Walter T. Federer, Biometrics Unit, Cornell University; John E. Freund, Department of Mathematics, Arizona State College; A. W. Wortham, Operations Research Department, Texas Instruments, Dallas, Texas.

The local staff includes: Carl E. Marshall (Ph.D., Iowa State), Franklin Graybill (Ph.D., Iowa State), Robert D. Morrison (Ph.D., North Carolina State, John Hamblen (Ph.D., Purdue), Roy Deal (Ph.D., University of Oklahoma), and John Hoffman (Ph.D., University of Oklahoma).

Of particular interest at this summer session will be the six weekly symposia covering six important areas in statistics. They are: Sampling Survey Designs, Experimental Designs, Non-parametric Statistics, Response Curves and Surfaces, Multiple Comparisons, and High Speed Computing. Discussants will be selected from major contributors to these areas. These invited speakers together with the outstanding summer school staff will cover the respective subjects from three points of view: applications, their mathematical bases, and the problems that lie on the frontier.

Inquiries should be addressed to Carl E. Marshall, Director, Statistical Laboratory, Oklahoma State University, Stillwater, Oklahoma.

List of International and Foreign Scientific and Technical Meetings

October 1, 1957 through 1960

(The following information was extracted from a list compiled by the Office of Scientific Information of the National Science Foundation.)

Date and Place	*Meeting, Sponsor and Subject*	*Address Queries to:*
Oct. 21, 1957 Lourenco Marques, Mozambique	2nd Inter-African Conference on STATISTICS, Inter-African Committee of Statistics	Secretariat, Commission for Technical Cooperation in Africa South of the Sahara, 43 Parliament Street, London, S. W. 1, England

Date and place	*Meeting, Sponsor and Subject*	*Address Queries to*
Dec. 26, 1957–Jan. 4, 1958 Berkeley, California	Symposium on AXIOMATIC METHOD, with Special Reference to Geometry and Physics	Professor Alfred Tarski, Department of Mathematics, University of California, Berkeley 4, California
Apr. 9–13, 1958 Giessen, Germany	Society for APPLIED MATHEMATICS AND MECHANICS (GAMM), Annual Meeting	Professor Dr. Egon Ullrich, Mathematisches Institut der Justus Liebig-Hochschule, Johannesstrasse 1, (16) Giessen, Germany
Apr. 13–16, 1958 Giessen, Germany	German MATHEMATICS SOCIETY (DMV), Annual Meeting	Professor Dr. Egon Ullrich, Mathematisches Institut der Justus Liebig-Hochschule, Johannesstrasse 1, (16) Giessen, Germany
June 1958 Strasbourg, France	International Association for *ANALOGY COMPUTATION*, 2nd International Meeting and 1st General Assembly	Professor J. Hoffmann, Université Libre, 50 Avenue Franklin D Roosevelt, Brussels, Belgium
Aug. 11–13, 1958 St. Andrews, Scotland	International MATHEMATICAL Union, 3rd General Assembly	Mr F Smithies, Mathematical Institute, 16 Chambers Street, Edinburgh 1, Scotland
Aug. 14–21, 1958 Edinburgh, Scotland	11th International Congress of MATHEMATICIANS—Logic and foundations, algebra and theory of numbers, analysis, topology; geometry, probability and statistics, applied mathematics, mathematical physics and numerical analysis; and history and education	Mr F. Smithies, Mathematical Institute, 16 Chambers Street, Edinburgh 1, Scotland
Sept. 3–10, 1958 Namur, Belgium	2nd International Congress for CYBERNETICS, Association Internationale de Cybernétique (ASBL)—Information—Automatism (Cybernetics applied to machinery)—Automation (Cybernetics used in organizing labor)—The economical and social consequences of Automation—Cybernetics and social sciences — Cybernetics and biology	International Association for Cybernetics, 13, rue Basse-Marcelle, Namur, Belgium
Sept. 1958 Brussels, Belgium	International STATISTICAL Institute, Special Session	Institut National de Statistique, 44, rue de Louvain, Brussels, Belgium
1958 Undecided Warsaw, Poland	International Symposium on NONHOMOGENEITY IN ELASTICITY AND PLASTICITY, International Union for Theoreti-	Dr Hugh L. Dryden, President of Union, NACA, 1512 H Street, N W., Washington, D C , or Professor F. N. van

Date and place	*Meeting, Sponsor and Subject*	*Address Queries to;*
	cal and Applied Mechanics (IUTAM)	den Dungen, Secretary of Union, 41 avenue de l'Arbalete, Boitsfont, Brussels, Belgium
1960 Undecided Stresa, Italy	10th International Congress of APPLIED MECHANICS, International Union of Theoretical and Applied Mechanics (IUTAM)	Dr. Hugh L. Dryden, President of Union, NACA, 1512 H Street, N. W., Washington, D. C.; or Professor F. N. van den Dungen, Secretary of Union, 41 avenue de l'Arbalete, Boitsfont, Brussels, Belgium

Royal Statistical Society Research and Industrial Applications Sections

The Research Section and the Industrial Applications Section of the Royal Statistical Society intend to hold a Conference at the University of St. Andrews, near Edinburgh, Scotland, from 22 August to 1 September inclusive. It will be devoted to Mathematical Statistics, with special reference to those branches of the subject which have application in industry.

It is proposed that there should be three morning sessions, consisting each of two or three pre-arranged lectures, and three early evening sessions (5.30 p.m. to 6.30 p.m.) each with one pre-arranged lecture.

The afternoons will be devoted to 'Splinter Groups' which will devote themselves to special aspects, and at which informal talks of some ten or fifteen minutes each can be given without prior arrangement.

Topics to be covered in the morning and evening sessions include aspects of the analysis of variance, non-parametric inference, stochastic aspects of linear and dynamic programming, and foundations of probability in statistics.

It is hoped that many of the mathematical statisticians who will be coming from abroad to attend the Edinburgh International Congress of Mathematicians, will choose to remain in Scotland for a further few days, and take the opportunity of meeting colleagues specially interested in their field. St. Andrews, besides having the famous Golf Course, is a small Scottish town of considerable character, and a very good centre for the exploration of the Eastern Highlands.

Accommodation (from 21 August to 2 September) will be provided within the hostels of the University of St. Andrews at a reasonably low cost, details to follow later. Anyone interested should write, marking the envelope 'ST. ANDREWS CONFERENCE' to Miss U. Croker, Royal Statistical Society, address as above.

University of Michigan Graduate School of Public Health Summer Program

The University of Michigan is the host institution for a cooperative program by the accredited Schools of Public Health of the United States during the summer of 1958.

The summer program for 1958 is designed to meet some of the educational and training needs of men and women engaged in work in health and health related agencies or those preparing themselves for such work. Courses are offered at three levels. The elementary level courses are intended for those who have acquired little or no background in statistical methodology. Intermediate courses present subject matter to extend and improve knowledge and skills of those persons who have acquired the elementary concepts and skills of statistical methodology. The advanced level courses are for those who have acquired considerable background in the theory and application of statistical concepts and procedures. A seminar open to all students includes the presentation of topics of current national interest related to health sciences and statistical methodology.

The faculty will consist of Helen Abbey, The Johns Hopkins University; William G. Cochran, Harvard University; Jerome Cornfield, The National Institutes of Health; Bernard Greenberg, University of North Carolina; F. M. Hemphill, University of Michigan, Leslie Kish, University of Michigan; Donovan Thompson, University of Pittsburgh; Colin White, Yale University.

If possible, completed applications and transcripts should reach Ann Arbor by June 1, 1958, for Michigan residents and May 1, for nonresidents. Requests for application forms should be addressed to the Director of the Summer Program in Public Health Statistics, School of Public Health, University of Michigan, Ann Arbor, Michigan.

A limited number of scholarships will be available to qualified students taking courses for credit. Inquiries concerning scholarships should be addressed to Dr. F. M. Hemphill, Director of the Summer Program in Public Health Statistics, School of Public Health, University of Michigan, Ann Arbor, Michigan.

IMS MEMBERS ATTENDING THE 1957 ANNUAL MEETING OF THE IMS

(This list was not received in time to be included with the report in the December, 1957 issue.)

Forman S Acton, Frank B. Akutowicz, William R Allen, Allan G. Anderson, R. L. [illegible] Anscombe, [illegible] A. Ban- [illegible] y Beall, Helen P. Beard, Robert Eric Bechofer, Charles Bernard Bell, Jr , Irving Belson, Andrew Angelo Benvenuto, Agnes P. Berger, Joseph Berkson, Gerald D. Berndt, Max A. Bershad, Reid A. T. Bhaucha, Charles A Bicking, Patrick Paul Billingsley, Richard S. Bingham, Jr., Allan Birnbaum, David Blackwell, Herman Blasbalg, Chester I Bliss, Julius R. Blum, Isadore Blumen, John B. Boddie, Derrill Joseph Bordelon, Raj C Bose, Helen Bozivich, Ralph Allan, Bradley, A. E. Brandt, Leroy S Brenna, Glenn W. Brier, Harold F. Bright, Samuel H. Brooks, Bernice Brown, Irwin D J Bross, Benjamin Buchbinder, Robert W. Burgess, Paul J. Burke, Irving W Burr, Glenn L Burrows, Lyle D Calvin, Burton H. Camp, C. S. Callard, Mavis B Carroll, Marvin F. Carter, Jack Chassan, Herman Chernoff, Victor Chew, Chin Long Chiang, John T. Chu, Joseph Louis Ciminera, Ira H. Cisin, Willard

H. Clatworthy, Andrew G. Clark, Charles William Clunies-Ross, William G. Cochran, Paul M. Cohen, William S. Connor, Clyde H. Coombs, Lewis C. Copeland, Richard G. Cornell, Jerome Cornfield, L. M. Court, Edwin L. Cox, Paul Charles Cox, Allen T. Craig, Elliot M. Cramer, Jean A. Crockett, Lee S. Crump, Edward Eugene Cureton, Joseph F. Daly, Cuthbert Daniel, Herbert Theodore David, Willis L. Davis, Read B. Dawson, Jr., W. Edwards Deming, Arthur P. Dempster, Cyrus Derman, Lucile Derrick, Earl Louis Diamond, John K. Diederichs, James L. Dolby, Tom G. Donnelly, Acheson J. Duncan, David B. Duncan, Paul R. Dunlap, Charles W. Dunnett, David Durand, Arthur Morlan Dutton, Meyer Dwass, Albert Ross, Eckler, Jr., Sylvain Ehrenfeld, Churchell Eisenhart, Harry Eisenpress, Salah A. Elmaghraby, Lila R. Elveback, Daniel R. Embody, Walter T. Federer, A. V. Fend, Robert Ferber, George Emery Ferris, William B. Fetters, Donald Fraser, David Frazier, Spencer M. Free, Spencer Michael Free, Jr., Agnes M. Galligan, Donald A. Gardiner, Werner Gautschi, Charles E. Gates, Donald Paul Gaver, Seymour Geisser, Lincoln J. Gerende, George William Gershefski, B. C. Getchell, Walter M. Gilbert, Dorothy Morrow Gilford, Leon Gilford, Harold Glazer, William A. Glenn, Ramanathan Gnanadesikan, Leo A. Goodman, Mina H. Gourary, Bernard G. Greenberg, Samuel W. Greenhouse, Joseph Arthur Greenwood, Frank E. Grubbs, Lee Gunlogson, John Gurland, Donald Guthrie, Irwin Guttman, Robert John Hader, Max Halperin, James F. Hannan, Morris Howard Hansen, Robert H. Hanson, Bernard Harris, T. E. Harris, Boyd Harshbarger, H. Leon Harter, Herman O. Hartley, William C. Healy, Jr., Paul Heit, F. M. Hemphill, G. Ronald Herd, Irene Hess, Clifford Hildreth, Wassily Hoeffding, Robert G. Hoffmann, John F. Hofmann, David Hogben, Paul G. Homeyer, Robert Hooke, Theodore Wright Horner, William H. Horton, Daniel G. Horvitz, Professor Harold Hotelling, Earl E. Houseman, David R. Howes, Walter W. Hoy, Cyril J. Hoyt, John David Hromi, Harry M. Hughes, J. Stuart Hunter, David V. Huntsberger, Benjamin Jackson, John L. Jaech, T. A. Jeeves, Milton Vernon Johns, Jr., Howard L. Jones, Hyman B. Kaitz, Samuel Karlin, Abraham, E. Karp, Marvin A. Kastenbaum, Leo Katz, Mort Keats, Oscar Kempthorne, Robert W. Kennard, George H. Kennedy, Bradford F. Kimball, Edgar P. King, Cal J. Kirchen, Leslie Kish, Truman L. Koehler, Martin Krakowski, Clyde Y. Kramer, William C. Krumbein, William Kruskal, Morton Kupperman, George M. Kuznets, Mrs. Helen Humes Lamale, Donald E. Lamphiear, Fred C. Leone, Howard Levene, Alfred Lieberman, Gerald J. Lieberman, Gilbert Lieberman, Jacob E. Lieberman, Julius Lieblein, Benjamin Lipstein, Stuart P. Lloyd, Frederic M. Lord, Eugene Lukacs, Bob Lundegard, George F. Lunger, John Hans MacKay, William G. Madow, Ralph A. Maggio, Clifford Joseph Maloney, Joseph Mandelson, Henry Berthold Mann, Eli S. Marks, Robert H. Matthias, Philip John McCarthy, Duncan C. McCune, Hartley Ellsworth McKean, Paul M. Meier, Margaret Merrell, W. Jay Merrill, Herbert A. Meyer, Paul D. Minton, Robert Mirsky, Sutton Monro, Alex M. Mood, Roger H. Moore, Donald Frank Morrison, Milton NMI Morrison, Norman Morse, Jack Moshman, Frederick Mosteller, Mervin E. Muller, R. B. Murphy, Jack Nadler, L. F. Nanni, Raymond Nassimbene, Joseph Anthony Navarro, August A. Carl Nelson, Jr., Peter E. Ney, S. I. Neuwirth, George E. Nicholson, Monroe L. Norden, Jack I. Northam, Horace W. Norton, Aloysius Joachim O'Connor, Junjiro Ogawa, Ingram Olkin, Paul S. Olmstead, Bernard Ostle, Donald B. Owen, William R. Pabst, Jr., Nancy S. Parker, Dr. Ellis F. Parmenter, Emanuel Parzen, John F. Pauls, Robert Nixon Pendergrass, B. E. Phillips, Eugene W. Pike, Edwin James George Pitman, Aloysius J. Polaneczky, Bruce P. Price, Ronald Pyke, Dana Edward Anthony Quade, Lila Knudsen Randolph, Herman Ravitch, Stanley Reiter, Elmer Edwin Remmenga, G. J. Resnikoff, William L. Roach, Jr., Spencer W. Roberts, Herbert Robbins, Douglas S. Robson, Robert Roeloffs, Harry M. Rosenblatt, Joan R. Rosenblatt, Murray Rosenblatt, Irving Roshwalb, S. N. Roy, Herman Rubin, David Rubinstein, Phillip Justin Rulon, Marion M. Sandomire, F. E. Satterthwaite, Sam Cundiff Saunders, Leonard S. Savage, Marvin A. Schneiderman, Seymour Max Selig, Richard H. Shaw, Sidney Shtulman, Walt R. Simmons, Rosedith Sitgreaves, John H. Smith, Thaddeus L. Smith, Jean E. Smolak, Milton Sobel, Herbert Solomon, Paul N. Somerville, Frederick

A. Sorensen, Melvin Dale Springer, John J Stansbrey, James Hall Stapleton, Roberg G. D. Steel, Arthur Stein, Frederick F. Stephan, Theodor D. Sterling, John N. Stewart, Ray B. Stiver, Jr., David S. Stoller, Samuel A Stouffer, Jacques St. Pierre, Hale C. Sweeny, Zen Szatrowski, Robert J. Taylor, James G C Templeton, Benjamin J. Tepping, Milton E. Terry, Earl A. Thomas, William Alfred Thompson, Jr, George W. Thomson, Leo J. Tick, John W. Tukey, Malcolm E. Turner, Hubertus Robert Van der Vaart, Herman W. Von-Guerard, *Helen M. Walker*, *David L.* Wallace, W Allen Wallis, John E. Walsh, Louis Weiner, Harry Weingarten, Irving Weiss, Phillips Whidden, Alfred G. Whitney, D. Ransom Whitney, John M. Wiesen, Frank Wilcoxon, Martin B Wilk, Samuel S Wilks, John W. Wilkinson, Evan James Williams, Gregory Williams, Myron J. Willis, Russell Lowell Wine, Gerald Winston, Max A. Woodbury, G Stanley Woodson, Charles Ashley Wright, Charles W. Wright, William J. Youden, Marvin Zelen, John Arthur Zoellner

Visiting Foreign Mathematicians

The following list of visiting foreign mathematicians has been received from the Division of Mathematics, National Academy of Sciences—National Research Council. The information given is, in order, the name, home country, host institution, and period of visit; AY stands for academic year 1957–1958.

Adams, John F.—U. K.—Institute for Advanced Study—AY, Adem, Jose—Mexico—Princeton University—Feb 1958–June 1958, Akizuki, Yasuo—Japan—University of Chicago—Oct. 1, 1957–June 30, 1958; Albertoni, Sergio—Italy—New York University—Sept. 1957–Feb. 1958; Andreotti, Aldo—Italy—Institute for Advanced Study (Sept 30, 1957–Dec. 20, 1957), Princeton University (Feb 1958–June 1958)—Sept 30, 1957–June 1958, Azumaya, Goro—Japan—Yale University—Sept. 1956–Sept 1958; Beale, E M L—U. K—Princeton University—Jan. 1958–Dec 1958; Birch, Bryan J—U. K—Princeton University—Sept. 1957–June 1958; Bjorck, Göran—Sweden—Institute for Advanced Study—AY; Bofinger, Victor J.—Australia—North Carolina State College—June 1957–April 1958; Burgers, Johannes M.—Netherlands—American University, National Bureau of Standards—Oct. 11, 1956–Oct. 1957; Carleson, Lennart—Sweden—Massachusetts Institute of Technology—Sept. 1957–Jan. 31, 1958; Cartier, Pierre—France—Institute for Advanced Study—AY; Chakravorti, J. G.—India—Brown University—AY; Chand, Uttam—India—Boston University—Jan. 1958–May 1958, Cohen, Daniel E—U K—Princeton University—Sept. 1957–June 1958; Copson, E T.—Scotland—Stanford University—Week-Feb 1958; Corsten, L. C. A.—Netherlands—University of North Carolina—Sept 1957–June 1958; Dedecker, Paul—Belgium—Institute for Advanced Study—Sept. 30, 1957–Dec 20, 1957; Delsarte, Jean—France—University of Maryland—Apr 1957–July 1957; Deny, Jacques—France—Institute for Advanced Study—Sept. 30, 1957–Dec 20, 1957; de Rham, Georges—Switzerland—Institute for Advanced Study—AY; Dold, Albrecht—Germany—Institute for Advanced Study—Sept. 1956–Aug. 1958; Dvoretzky, Aryeh—Israel—Institute for Advanced Study—AY; Edwards, David A.—U. K—Yale University—Sept. 1956–Sept. 1958; Ewald, Guenther—Germany—Michigan State University—Sept 1957–June 1958; Festa, Rudolf—Austria—University of Alabama, 1956–57 (State College of Washington 1957–58) —Sept. 1956–Sept. 1958; Festa, Erika—Austria—State College of Washington (Sept.–Dec. 1957)—Sept. 1956–Sept. 1958; Foguel, Shaul—Israel—New York University—AY; Fröhlich, A.—U. K.—University of Virginia—Feb. 1958–June 1958; Gamblen, Frank—Australia—University of Kansas (Sept. 4, 1957–Feb. 1, 1958), Educational Testing Service, Princeton, N. J. (Feb. 1, 1958–June 1958)—Sept. 4, 1957–June 1958; Gautschi, Walter—Switzerland—American University, National Bureau of Standards—Oct. 1955–Sept. 1958; Ghaffari, A.

—Iran—American University, National Bureau of Standards—Sept. 1956–Sept. 1958; Goldner, Siegfried—Union of South Africa—New York University—AY; Grauert, Hans—Germany—Institute for Advanced Study—AY; Grenander, U.—Sweden—Brown University—AY; Griffiths, Hubert B.—U. K.—Institute for Advanced Study—Sept. 1956–July 1958; Guttman, Irwin—Canada—Princeton University—Sept. 1957–June 1958; Hano, Jun-ichi—Japan—University of Washington—Sept. 15, 1957–June 15, 1958; Harrop, Ronald—U. K.—Pennsylvania State University—Aug. 1957–Aug. 1958; Hellman, Olavi B.—Finland—University of California, Los Angeles—July 1956–July 1958; Helmberg, Gilbert—Austria—University of Washington—October 1, 1957–June 15, 1958; Hervé, Michel—France—Institute for Advanced Study—Sept. 30, 1957–Dec. 20, 1957; Hirsch, Guy—Belgium—Massachusetts Institute of Technology—Feb. 1, 1958–June 15, 1958; Hitotumatu, Sin—Japan—Stanford University—Sept. 1, 1957–June 30, 1958; Hüsser, Rudolph—Switzerland—University of California, Los Angeles, AY; Izumi, Shin-ichi—Japan—University of Chicago and Northwestern University (Aug. 1, 1957-May 31, 1958), Princeton University (Oct. 1957–Dec. 1957)—Aug. 1957–May 1958; Kato, Tosio—Japan—New York University—Sept.-Oct. 1957; Kawata, T.—Japan, Princeton University—Sept. 1957–March 1958; Kitawaga, T.—Japan—Princeton University—Sept. 1957–March 1958; Klingenberg, Wilhelm—Germany—Institute for Advanced Study—AY; Laasonen, V. Pentti J.—Finland—University of California, Los Angeles—July 1956–Aug. 1958; Lacombe, Daniel L. M.—France—Institute for Advanced Study—Oct. 1957–Aug. 1958; Lehto, Olli E.—Finland—Institute for Advanced Study—AY; Leopoldt, Heinrich W.—Germany—Institute for Advanced Study—Sept. 1956–Aug. 1958; Leray, Jean—France—Institute for Advanced Study—Sept. 30, 1957–Dec. 20, 1957; Lions, Jacques—France—University of Kansas—Feb. 1957–Aug. 1958; Longuet-Higgins, Michael S.—U. K.—Massachusetts Institute of Technology—Feb. 1, 1958–June 15, 1958; Lorenzen, Paul P. W.—Germany—Institute for Advanced Study—Sept. 1957–June 1958; Lucas, John R.—U. K.—Princeton University—Sept. 1957–June 1958; Lumer, Günter—Uruguay, University of Chicago—Oct. 1, 1957–Sept. 30, 1958; Mallows, Colin L.—U. K.—Princeton University—Sept. 1957–Sept. 1958; Mardešić, Sibe—Jugoslavia—Institute for Advanced Study—AY; Martin, Alfred I. —U. K.—Institute for Advanced Study—AY; Masani, Pesi—India—Massachusetts Institute of Technology—Harvard University—Sept. 16, 1957–June 15, 1958; Message, Philip J.—U. K.—Yale University—Sept. 1957–Sept. 1958; Milne-Thomson, L. M.—U. K.—Brown University—Sept. 1956–June 1958; Mixner, Joseph—Germany—New York University—Aug.-Oct. 1957; Møller, Christian—Denmark—Carnegie Institute of Technology—Sept. 1957–Feb. 1958; Nachbin, Leopoldo—Brazil—University of Chicago—Oct. 1, 1956–July 31, 1958; Nagata, Masayoshi—Japan—Harvard University—Sept. 1957–Sept. 1958; Nieminen, Toivo E.—Finland—New York University—Aug. 1957–June 1958; Ogawa, Junjiro—Japan—Institute of Statistics, University of North Carolina—Sept. 1956–Aug. 31, 1958; Olver, F. W. J.—U. K.—National Bureau of Standards—Sept. 30, 1957–Sept. 1958; O'Meara, Onorato T.—South Africa—Institute for Advanced Study—AY; Ono, Katsuzi—Japan—Massachusetts Institute of Technology—Sept. 1957–June 1958; Ostrowski, Alexander M.—Switzerland—American University, National Bureau of Standards—Oct. 10–31, 1957; Papakyriakopoulos, C. D.—Greece—Institute for Advanced Study—June 3, 1955–June 1958; Peixoto, Mauricio M.—Brazil—Princeton University—Sept. 1957–June 1958; Pfluger, Albert—Switzerland—Stanford University—Oct. 1, 1957–Mar. 30, 1958; Pucci, Carlo—Italy—University of Maryland—Sept. 1, 1956–July 1958; Puppe, Dieter—Germany—Institute for Advanced Study—Sept. 1957–June 1958; Rieger, Georg J.—Germany—University of Maryland—Sept. 1956–Aug. 1957; Riesz, Marcel—Sweden—University of Maryland—Oct. 1, 1957–Dec. 31, 1957; Robinson, Leslie R. B.—U. K.—Harvard University—Sept. 1957–Sept. 1953; Rogosinski, Werner W.—U. K.—University of Colorado—Sept. 1957–Sept. 1958; Rohrbach, Hans—Germany—University of North Carolina—Sept. 1957–June 1958; Room, T. G.—Australia—Institute for Advanced Study (AY); Princeton University (Feb. 1958–June 1958)—Sept. 1957–June 1958; Roseau, Maurice—France—New York University

—Sept. 1957-Sept. 1958; Sawyer, W. W.—U. K.—University of Illinois—Feb. 1957-Indefinite; Schopf, Andreas—Switzerland—American University-National Bureau of Standards—Sept. 30, 1957-Oct. 1958; Scriba, J Cristoph—Germany—University of Kentucky—Sept. 1957-June 1958; Selberg, Sigmund—Norway—University of Colorado—Sept. 1957-May 1958; Serre, Jean-Pierre—France—Institute for Advanced Study—Sept. 20, 1957-Dec. 20, 1957; Skolem, Thoralf A —Norway—University of Notre Dame—Sept. 1957-June 1958; Stoll, Wilhelm—Germany—Institute for Advanced Study—AY; Tamagawa, Tsuneo—Japan—Institute for Advanced Study (Sept. 1955-Jan. 1957 and Jan. 13, 1958-Apr. 11, 1958), Johns Hopkins University (Jan. 1957-Jan 1958)—Sept. 1955-Apr 1958; Tomonaga, Yasuro—Japan—University of Washington—Oct 15, 1957-June 15, 1958; Valpola, Veli Kustaa—Finland—University of California (2 months), Princeton University (4 months)—Nov. 1957-April 1958; van der Vaart, H R —Netherlands—North Carolina State College—Jan. 1957-Jan. 1958; Villamayor, Orlando—Argentina—Institute for Advanced Study—Jan. 1, 1957-Dec. 31, 1957, Waelbroeck, L —Belgium—Institute for Advanced Study—AY; Watson, Geoffrey S —Australia—North Carolina State College (4 months), Princeton University (5 months)—April 1958-Dec 1958, Williams, Robert M —U K —Princeton University—Sept 1957-June 1958, Wolff, Emil—U K —New York University—June-Nov. 1957; Yamamuro, Sadayuki—Japan—Institute for Advanced Study—Sept 1956-Sept 1958; Yevdjevich, V. M.—Yugoslavia—American University, National Bureau of Standards—AY; Yüksel, H.—Turkey—Brown University—AY, Zadunaisky, Pedro—Argentina—Watson Scientific Computing Laboratory—Feb 1, 1957-Jan. 31, 1958; Agudo, F R. D —Portugal—University of California, Berkeley Fall 1957; Baayen, Pieter C —Netherlands—University of California, Berkeley—AY; Fary, Istvan—Canada (Hungary)—University of California, Berkeley Jan -June 1958, Festa, Erika—Austria—State College of Washington Sept.-Dec. 1957; Lightstone, A H.—Canada—University of California, Berkeley—AY; Littlewood, J. E —England—University of California, Berkeley Sept. 27-Dec 18, 1957; Poulsen, Ebbe T.—Denmark—University of California, Berkeley July 1957-June 1958; Specker, Ernst Paul—Switzerland—Cornell University Feb 10-Sept 1958; Szmielew, Wanda—Poland—University of California, Berkeley—AY.

Committee on Mathematical Tables

Since its organization early in 1956, the Institute of Mathematical Statistics' Committee on Mathematical Tables has been concerned with the problems associated, either directly or indirectly, with the computation of mathematical tables of interest to statisticians. The committee's function is threefold:

(i) To gather information relating to the tabulation of functions of interest to statisticians.

(ii) To advise on the need for, and preparation of, statistical tables.

(iii) To determine the availability of and coordinate the distribution of free time on high speed digital computers for the computation of statistical tables.

In order to fulfill its function, the committee investigated the interests and needs of the Institute membership concerning statistical tables and as a result set up nine subcommittees covering the areas of greatest interest. To date the activities of these subcommittees have been directed primarily towards the preparation of bibliographies in their individual fields. The subcommittees, along with their chairmen, are listed below.

1. Chi-Square,
   W. Kruskal, University of Chicago
2. t-Distribution (Univariate and Multivariate),
   C. W. Dunnett, American Cyanamid Company, Pearl River, New York.
3. Studentized Range,
   A. H. Bowker, Stanford University
4. F-Distribution (Incomplete-Beta, Binomial),
   E. E. Cureton, University of Tennessee.
5. Hypergeometric Distribution (*Not* the hypergeometric function),
   W. Kruskal, University of Chicago.
6. Polyvariate Normal Distribution, including latent roots,
   G. P. Steck, Sandia Corporation, Albuquerque, New Mexico
7. Availability of Simple Techniques,
   Chairman not appointed.
8. Annals Supplement (Of statistical tables),
   J. W. Tukey, Bell Telephone Laboratories, Murray Hill, New Jersey
9. Computing Facilities and Cost-Free Machine Time,
   F. C. Leone, Case Inst. of Tech.

Additional information on any of the above activities may be obtained from the chairman of the Committee on Mathematical Tables, D. B. Owen, Sandia Corporation, Albuquerque, New Mexico, or from any of the subcommittee chairmen. Anyone having time on a digital computer which may be made available on a cost-free basis to persons desiring to compute tables of general interest is invited to contact the chairman of subcommittee 9.

I. *List of Members of the IMS Committee on Mathematical Tables*

Chairman
Dr. D. B. Owen, Division 5125, Sandia Corporation, Albuquerque, New Mexico

Secretary and Vice Chairman
Dr. G. P. Steck, Division 5125, Sandia Corporation, Albuquerque, New Mexico

Members
Professor R. L. Anderson, Institute of Statistics, North Carolina State College, Raleigh, North Carolina
Professor A. H. Bowker, Department of Statistics, Stanford University, Stanford, California
Professor E. E. Cureton, 1846 Prospect Pl., S.E., Knoxville 15, Tennessee
Professor W. J. Dixon, University of California, Department of Preventive Medicine and Public Health, Medical Center, Los Angeles 24, California
Mr. C. W. Dunnett, 19 Edsall Place, Nanuet, New York
Dr. Churchill Eisenhart, Chief, Statistical Engineering Laboratory, National Bureau of Standards, Washington 25, D. C.
Dr. J. A. Greenwood, 16 Garfield Street, Cambridge 38, Massachusetts
Professor H. O. Hartley, Statistical Laboratory, Iowa State College, Ames, Iowa
Professor William Kruskal, Committee on Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois

Professor Fred C. Leone, Director, Statistical Laboratory, Case Institute of Technology, 10900 Euclid Avenue, Cleveland 24, Ohio
Professor Dan Teichroew, Graduate School of Business, Stanford University, Stanford, California
Dr. John W. Tukey, Bell Telephone Laboratories, Murray Hill, New Jersey
Professor M. A. Woodbury, 401 W. 205th Street, New York 34, New York
Dr. Marvin Zelen, Statistical Engineering Laboratory, National Bureau of Standards, Washington 25, D. C.

Ex officio members

Professor L. J. Savage, Eckhart Hall, University of Chicago, Chicago 37, Illinois
Professor Jacob Wolfowitz, Dept of Math, Cornell University, Ithaca, New York

II. *Subcommittees of the IMS Committee on Mathematical Tables*

1. *Chi-square*
   *William Kruskal
   *Dan Teichroew

2. *t-distributions* (univariate and multivariate)
   **C. W. Dunnett, Chairman*
   Dr. H. A. David, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia
   Dr. S. S. Gupta, Department of Mathematics, University of Alberta Edmonton, Alberta, Canada
   *H. O. Hartley
   Professor E. S. Keeping, Math Department, University of Alberta, Edmonton, Alberta, Canada
   Professor C. F. Kossack, Math Department, Purdue University, West Lafayette, Indiana
   Dr. A. M. Mood, General Analysis Corporation, 11753 Wilshire Boulevard, West Los Angeles, California
   J. B. Rabin, Sen. Computer Analyst, Burroughs Corporation, 1505 Sycamore Avenue, Willow Grove, Pennsylvania
   Dr. M. Sobel, Bell Telephone Lab, 555 Union Boulevard, Allentown, Pa
   *Dan Teichroew

3. *Studentized range*
   *A. H. Bowker, Chairman,
   Cuthbert Daniel, 116 Pinehurst Avenue, New York 33, New York
   Prof. W. T. Federer, Cornell University, Ithaca, New York
   *H. O. Hartley
   Professor G. E. Noether, Math Department, Boston University, 725 Commonwealth Ave., Boston 15, Massachusetts

4. *F-distribution* (incomplete-beta, binomial)
   *E. E. Cureton, Chairman,
   *R. L. Anderson
   P. C. Cox, 1904 Idaho Avenue, Las Cruces, New Mexico
   Professor David Durand, 50 Memorial Drive, Cambridge 39, Massachusetts
   *J. A. Greenwood
   *H. O. Hartley
   Gunnar Kulldorff, University of Lund, Lund, Sweden, Malmgatan 16, Malmo, Sweden
   *Dan Teichroew

* Member of the parent committee See List I for address

**H. F. Trotter
Prof. J. G. Wendel, Math Department, University of Michigan, Ann Arbor, Michigan

5. *Hypergeometric distribution* (*not* the hypergeometric function)
*William Kruskal, Chairman,
Professor Leo Katz, Department of Statistics, Michigan State University, East Lansing, Michigan
Dr. E. F. Kimball, N. Y. State Pub. Service Commission; 20 Mayfair Drive, Slingerlands, New York
Professor G. J. Lieberman, Department of Statistics, Stanford University, Stanford, California
Roger H. Moore, Los Alamos Scientific Laboratory, 3448A Orange, Los Alamos, New Mexico
J. M. Wiesen, 1308 Arizona NE, Albuquerque, New Mexico

6. *Polyvariate normal*, including latent roots
*G. P. Steck, Chairman
Professor T. W. Anderson, Center for Advanced Study in the Behavioral Sciences, 202 Junipero Serra Blvd., Stanford, California
P. C. Cox, (See Subcommittee 4 for address)
*C. W. Dunnett
S. S. Gupta, (See Subcommittee 2 for address)
Professor Ingram Olkin, Department of Statistics, Michigan State University, East Lansing, Michigan
*D. B. Owen
M. Sobel, (See Subcommittee 2 for address)
*Max A Woodbury

7. *Availability of simple techniques*
Chairman position open.
Professor R. A. Bradley, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia
*E. E. Cureton
*W. J. Dixon
*Churchill Eisenhart
Dr. T. A. Lamke, Bu. of Res., Iowa State Teachers College, Cedar Falls, Iowa
Professor S. B. Littauer, Columbia University, New York 27, New York
**H. R. Watkins

8. *Annals Supplement*
*J. W. Tukey, Chairman
J. S. Barnes, John Wiley & Sons Inc., 440 Fourth Avenue, New York 16, N. Y.
*A. H. Bowker
*Churchill Eisenhart
Dr. T. E. Harris, The RAND Corporation, 1700 Main Street, Santa Monica, Calif.
*D. B. Owen
*Dan Teichroew

9. *Cost-Free Machine Time*
*F. C. Leone, Chairman
Professor J. W. Hamblen, Computing Center, Oklahoma State University, Stillwater, Oklahoma
W. H. Horton, Materials Engineering Department, Westinghouse Electric Corp., East Pittsburgh, Pa.

** Address unknown.

G. F. Lunger, Program Planning Department, Remington Rand Univac, St. Paul 16, Minnesota
Dr. H. A. Meyer, Director Statistical Laboratory, University of Florida, Gainesville, Florida
*Max A. Woodbury

REPORT OF THE LOS ANGELES MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The Western Region Meeting, seventy-fifth meeting of the Institute of Mathematical Statistics, was held at the Los Angeles Campus of the University of California on December 27–28, 1957. Two sessions were joint with the American Statistical Association and the Institute for Management Sciences. A Special Invited Address was given by Claude Shannon, "Some Asymptotic Estimates for Sums of Random Variables." The following 69 members of the Institute attended:

I. J. Abrams, T. W. Anderson, H. L. Ang, Leo A. Aroian, C. B. Bell, D. L. Bentley, Allan Birnbaum, Colin R. Blyth, Charles Boll, Julien L. Borden, A. H. Bowker, J. V. Breakwell, Bernice Brown, Herman Chernoff, Edward P. Coleman, L. M. Court, Edwin L. Crow, W. J. Dixon, Olive Jean Dunn, H. P. Edmundson, Bob E. Ellison, T. S. Ferguson, Evelyn Fix, Martin Fox, A. V. Gafarian, Edward Gammon, Norman R. Garner, E. J. Gilbert, F. A. Graybill, Wm C. Guenther, D. Guthrie, Jr., T. E. Harris, P. G. Hoel, John P. Hofmann, John M. Howell, Arnljot Høyland, Patricia Inman, M. V. Johns, Jr., R. F. Link, Albert Madansky, Craig A. Magwire, F. Massey, M. R. Mickey, O. B. Moan, Roger A. Moore, Paul B. Moranda, James Pachares, Emanuel Parzen, M. P. Peisakoff, H. H. Peterson, Ron Pyke, Roy Radner, F. C. Reed, David Rothman, Marion M. Sandomire, Henry Scheffé, E. M. Scheuer, Franklin Sheehan, Bernard Sherman, M. M. Siddiqui, Paul N. Somerville, D. Stoller, Fred H. Tingey, Howard G. Tucker, John W. Tukey, H. W. von Guérard, John E. Walsh, Louis H. Wegner, Bryan Wilkinson

The program of the meeting was as follows:

Friday, December 27, 1957

8:45–12:00 a.m. Statistics in the Management Sciences
(Joint Session with The Institute of Management Sciences)

Chairman: M. R. Mickey, Jr., The RAND Corporation.

1. *The Portfolio Selection Problem*, Harry Markowitz, The RAND Corporation.
2. *On the Stochastic Theory of Inventory*, I. J. Abrams, The Ramo-Wooldridge Corporation.
3. *Inventory Control Problems of Shipboard Supplies*, Mina H. Gourary, George Washington University. (Read by Bernice B. Brown, The RAND Corporation).
4. *Demand for and Allocation of Engineering Personnel*, Rajendra Ka[illegible] (introduced by H. W. von Guerard) and Herman W. von Guerard, Lockheed Aircraft Corporation.

1:45–2:15 p.m. Special Invited Paper

Chairman: Thomas S. Ferguson, University of California, Los Angeles
Some Asymptotic Estimates for Sums of Random Variables
Claude Shannon, MIT, The Center for Advanced Study in the Behavorial Sciences, and Bell Telephone Laboratories.

2:30–5:00 p.m. Industrial Applications of Statistics

(Joint Session with the American Statistical Association)
Chairman: John F. Hofmann, Systems Laboratories Corporation

1. *Some Applications of Experimental Design in Industry*, Alex M. Mood and Paul Somerville, General Analysis Corporation.
2. *The Fitting of a Polynomial Form to a Function of Several Variables by the Use of Orthogonal Latin Squares*, N. M. Peterson, Convair, Fort Worth.
3. *Confidence Intervals for the Reliability of Multi-Stage Systems*, William C. Hoffman, The RAND Corporation.
4. *A Model for Depicting Fatigue*, Irvin Whiteman, General Analysis Corporation (introduced by A. M. Mood).
5. *Long Range Planning for Manufacturing*, Glen Ghormley, Cannon Electric Company.

Saturday, December 28, 1957

8:45–10:45 a.m. Stochastic Processes Applied to Medicine and Public Health

Chairman: Frank J. Massey, University of California, Los Angeles

1. *Replication Versus Increasing Observation Points in the Estimation of Regression for Growth Type Data*, Paul G. Hoel, University of California, Los Angeles.
2. *The Identifiability Problem for Functions of Finite Markov Chains*, Edgar John Gilbert, University of California, Berkeley.

11:00–12:00 a.m. Invited Address

Chairman: Roger A. Moore, The Ramo-Wooldridge Corporation

1. *Statistical Theory of Some Quantal Response Models*, Allen Birnbaum, Columbia University.

1:45–2:45 p.m. Invited Address

Chairman: O. B. Moan, Lockheed Aircraft Corporation

1. *Experiments with Mixtures*, Henry Scheffé, University of California, Berkeley.

3:00–5:00 p.m. Contributed Papers

Chairman: Richard F. Link, Oregon State College

1. *Non-parametric Multiple-decision Procedures for Selecting That One of K Populations which has the Highest Probability of Yielding the Largest Obser-*

vation. (Preliminary Report). Robert Bechhofer, Cornell University, and Milton Sobel, Bell Telephone Laboratories. (By Title)

2. *The Asymptotic Efficiency of Friedman's Chi-square$_r$-test* (χ_r^2-test). Ph. van Elteren, Mathematical Centre, Amsterdam. (By title)
3. *Least-squares Estimation when Residuals are Correlated.* M. M. Siddiqui, University of North Carolina.
4. *A Property of Additively Closed Families of Distributions.* Edwin L. Crow, Boulder Laboratories, National Bureau of Standards
5. *Determining Sample Size for a Specified Width Confidence Interval* Franklin A. Graybill, Oklahoma State University.
6. *Nonparametric Estimation of Sample Percentage Point Standard Deviation.* John E. Walsh, Lockheed Aircraft Corporation.
7. *On the Structure of Distribution-free Statistics* C. B. Bell, Xavier University of Louisiana and Stanford University.
8. *Estimation of the Location of a Discontinuity in Density.* J. V Breakwell, Lockheed Missile Systems Division, Palo Alto, and H Chernoff, Stanford University.
9. *On the Supremum of the Poisson Process.* Ronald Pyke, Stanford University.

EVELYN FIX
Associate Secretary

REPORT OF THE EDITOR OF THE ANNALS FOR 1957

During the year ending August 1, 1957, more new manuscripts were received by the *Annals*, totaling more manuscript pages, than in any previous year. A consequent increased requirement for printing is anticipated for the coming year, and the Council has authorized a 1958 volume of 1300 pages

The 1957 volume, totaling 1098 pages, contained 105 papers and notes The increased size authorized by the Council in the past two years made it possible to keep the backlog at less than half an issue during 1957.

The *Annals* is indebted to its staff of Cooperating Members, who do much of the refereeing, and to the following people who have generously given refereeing assistance: T. W. Anderson, P. Armitage, E. W. Barankin, M. S. Bartlett, R. Blumenthal, R. Bechhofer, G. E. P. Box, L. Breiman, D. L. Burkholder, S Chandrasekhar, W. G. Cochran, W. S. Connor, L. Cote, S. L. Crump, F. N. David, M. D. Donsker, R. Dorfman, A. Duncan, M. Dwass, B. Epstein, P. Erdős, T. S. Ferguson, L. J. Folks, E. J. Gilbert, I. J. Good, L. Goodman, F. Graybill, U. Grenander, S. S. Gupta, J. F. Hannan, M. H. Hansen, W. Hoeffding, P. G. Hoel, H. Hotelling, A. T. James, N. L. Johnson, E. S. Keeping, J. H. B. Kemperman, D. G. Kendall, M. G. Kendall, H. Kesten, E. L. Lehmann, J. Leiblein, R. Leipnik, M. Loève, E. Lukacs, J. McGregor, A. Madansky, W. G. Madow, M. R. Mickey, A. M. Mood, P. A. P. Moran, F. Mosteller, J. Moyal, R. W. Murphy, M. Newman, I. Olkin, E. Parzen, R. L. Plackett, J. Pratt,

R. Pyke, C. R. Rao, D. Ray, G. E. H. Reuter, J. Riordan, M. Rosenblatt, H. L. Royden, H. Rubin, J. Sacks, I. R. Savage, H. Scheffé, E. L. Scott, J. F. Scott, R. Sitgreaves, C. Streibel, R. F. Tate, D. Teichroew, A. J. Thomasian, H. Trotter, J. W. Tukey, D. L. Wallace, J. E. Walsh, B. L. Welch, L. Wegner, R. A. Wijsman, D. M. G. Wishart, G. Zyskind.

Many thanks are due Ann Greene, Dorothy Stewart, and Margaret Wray, for handling the taxing work of the editorial office.

T. E. Harris
Editor

December 26, 1957

SUMMER SESSIONS AT BERKELEY, CALIFORNIA

The 1958 summer program in the Department of Statistics of the University of California, Berkeley, California, will consist of two sessions: June 16 to July 26 and July 28 to September 6. The faculty of the summer sessions will include Professor U. S. Nair of Travancore University in India, Dr. F. N. David of University College in London, and Professors David Blackwell, Evelyn Fix, Joseph L. Hodges, Jr., and J. Neyman of the Department of Statistics of the University of California, Berkeley. The program will include two undergraduate courses in each session, and two research seminars, one in statistical problems of health and one in the statistical study of structural relations in the physical sciences.

PUBLICATIONS RECEIVED

Chow, G. C., *Demand for Automobiles in the United States*, North-Holland Publishing Company, Amsterdam, (1957) v-110 pp., $10.00.

Economica, *Revista de la Facultad de Ciencias Economicas*, Publicacion Trimestral, Diagonal 77-4 Y 5, La Plata, Buenos Aires, Argentina.

Contributions to the Theory of Games, Volume III, Edited by M. Dresher, A. W. Tucker, P. Wolfe, Princeton, New Jersey, Princeton University Press, 1957 Annals of Mathematics Studies Number 39.

INFORMATION THEORY FOR MATHEMATICIANS[1]

By J. Wolfowitz

Cornell University

A more descriptive term for information theory and one preferred by the present writer is "the theory of coding of messages." In this expository note we will describe briefly some basic concepts of this theory when transmission is through a "noisy channel" (noise = chance errors). We shall assume that both the transmitting alphabet and the receiving alphabet consist of two symbols, 0 and 1, say. This represents no loss in generality because the extension to any other alphabet, say one of twenty-six symbols, is immediate and presents no difficulty at all.

The fundamental paper of the theory is [1], other important papers are [2], [3], [4], and [5]. The papers most easily intelligible to the mathematician are probably [3], [4], [7], and [8]. The latter three deal with the subject matter of the present paper; [4] and [7] may each be read without any prior reading, and [8] is a sequel of [7]. In the present paper we describe four theorems proved in [7] and [8] and their relation to prior results.

Suppose that a person has a vocabulary of S words, any of which he may want to transmit, in any frequency and in any order, over some channel. We emphasize that we do not assume anything about the frequency with which particular words are transmitted, nor that the words to be transmitted are selected by any random process; in this respect our treatment differs from most of those in the literature.

Let the words be numbered in some fixed but arbitrary manner. Then transmitting a word is equivalent to transmitting one of the integers $1, 2, \cdots, S$. Let $s = \log S$ (all logarithms in this paper are to the base 2). Then there are S sequences of s elements each[2], each element either 0 or 1. If there is no noise, i.e., error of transmission, then, to transmit any word one has only to transmit the appropriate sequence of s zeros or ones.

If there is noise then this is clearly not enough, for the transmitted sequence will usually be incorrectly received. What is needed is that the received sequence, which will usually be a moderately garbled version of the transmitted sequence, should still be different from the moderately garbled version of any other transmitted sequence, so that one can infer what sequence it is that has been transmitted. But this requires that the sequences to be sent be not too similar in some reasonable sense, lest they be confused in transmission. Hence one must employ sequences of length greater than s, and *not all* such sequences (so that "neighboring" sequences be not sent). All these remarks will now be made precise.

Let the integer m ($\geqq 0$) be the "memory". A sequence of n (respectively $(n - m)$,

Received October 31, 1957.

[1] Rietz lecture delivered (under a different title) at the Atlantic City meeting of the Institute of Mathematical Statistics on September 10, 1957, by invitation of the Council of the Institute. Work under contract with the Office of Naval Research.

[2] Obviously, if s is not an integer one should replace it by the smallest integer $\geqq s$.

$(m + 1))$ elements, each zero or one, will be called an x-sequence (resp., a y-sequence, an α-sequence).[3] A transmitted sequence (received sequence) is always an x-sequence (y-sequence). There is given a "channel probability function" p, defined in the domain of all α-sequences, such that, for any α-sequence α, $0 \leqq p(\alpha) \leqq 1$. The "noisy channel" transmits an x-sequence x as follows: Let α_1 be the α-sequence of the first $(m + 1)$ elements of x. The channel "performs" a chance experiment with one of two possible outcomes, 1 and 0, with respective probabilities $p(\alpha_1)$ and $1 - p(\alpha_1)$. The outcome of the experiment is the first element of the received sequence $Y(x)$. Let α_2 be the α-sequence of the 2nd, 3rd, $\cdots$, $(m + 2)$th elements of x. The channel now performs a chance experiment, independent of the first, with possible outcomes 1 and 0 and respective probabilities $p(\alpha_2)$ and $1 - p(\alpha_2)$. This is repeated until $(n - m)$ independent experiments have been performed. The probabilities of the outcomes one and zero in the ith experiment are $p(\alpha_i)$ and $1 - p(\alpha_i)$, respectively, where α_i is the α-sequence of the ith, $(i + 1)$st, $\cdots$, $(i + m)$th elements of x. The received sequence $Y(x)$ is a *chance* y-sequence made up of the outcomes of the experiments in consecutive order. Let y_1 be any y-sequence. If $P\{Y(x) = y_1\} > 0$ (the symbol $P\{\ \}$ denotes the probability of the relation in braces) then y_1 is called a possible received sequence when x is transmitted.

Let λ, $0 < \lambda < 1$, be a given number. A "code" of length t is a set $\{(x_i, A_i), i = 1, \cdots, t\}$ where each x_i is an x-sequence, each A_i is a set of y-sequences, the A_i are all disjoint, and for each i, $i = 1, \cdots, t$,

$$P\{Y(x_i)\ \varepsilon\ A_i\} \geqq 1 - \lambda.$$

To be able to transmit S words we need a code of length S. The practical application of a code is as follows: When one wishes to transmit the ith word one transmits the x-sequence x_i. Whenever the receiver receives a y-sequence which is in A_j, he always concludes that the jth word has been sent. When the receiver receives a y-sequence not in $A_1 \cup A_2 \cup \cdots \cup A_t$ he may draw any conclusion he wishes about the word that has been sent. The probability that any word transmitted will be incorrectly received is $< \lambda$.

The quantity $(1/n) \log t$ is called the rate of transmission. The practical advantages of a high rate of transmission are obvious. In this paper we shall be concerned with the problem of determining, or at least bounding, the highest possible rate of transmission.

If $p(\alpha_1) = p(\alpha_2)$, then the two α-sequences α_1 and α_2 are indistinguishable in transmission. Barring such cases for simplicity, then, whatever be λ, $0 < \lambda < 1$, it is always possible to find an n and then a code of length S, provided one is willing to transmit at a sufficiently small rate. By sufficient repetition of the word to be transmitted one can insure that the probability of its correct reception exceeds $1 - \lambda$.[4]

---

[3] These terms are used only in [7] and [8].

[4] For example, "estimation" of the word transmitted may be by the method of maximum likelihood. The words in the vocabulary are the possible "values" of the parameter to be estimated. Since there are only finitely many words in the vocabulary the method of maximum likelihood is uniformly consistent.

What we have called a code in the present paper is usually called "an error correcting" code[5] in the literature of coding theory. The latter often admits as codes systems which do not meet the definition of code given above. Much of the literature of coding theory is concerned with the situation where the words to be transmitted are chosen from the vocabulary by a chance process with known distribution. Without discussing this matter further here we invite the reader to verify that the results cited below about the existence of (error correcting) codes of certain lengths hold a fortiori when the words to be transmitted are chosen by a chance process.

Let M_2 be the class of all stationary, metrically transitive stochastic processes

$$X_1, X_2, X_3, \cdots$$

where the chance variables X_i can take only the values 0 and 1 Let M_1 be the subclass of M_2 in which the X_i constitute a Markov chain. Let M_0 be the subclass of M_1 in which the X_i are independently distributed. We shall shortly define a functional φ on every member of M_2 (more precisely, φ will be a functional of the distribution functions of the stochastic processes). In the meantime, let C_2, C_1, C_0, be, respectively, the supremum of φ over M_2, M_1, M_0, respectively Then, of course, $C_0 \leqq C_1 \leqq C_2$.

Let ϵ always be an arbitrary positive number. The following Theorem A was first proved by Shannon [1] for the situation when the words to be transmitted are chosen by a known random process and for, in general, not error correcting codes.

THEOREM A. *For sufficiently large n there exists a code of length*

$$2^{n(C_1-\epsilon)}$$

(In [1] Shannon not only proved this remarkable theorem but brilliantly laid the foundations of the whole subject). Basing himself on the ingenious and important work of Feinstein [2] and McMillan [3], Khintchine in a very important paper [4] rigorously proved THEOREM B. *For sufficiently large n there exists a code of length*

$$2^{n(C_2-\epsilon)}.$$

While Khintchine's paper does not explicitly treat error correcting codes, one can deduce from his proof that Theorem B holds for error correcting codes.

Theorem B obviously implies Theorem A (both for error correcting codes). The question arises whether Theorem B is stronger than Theorem A, i.e., whether $C_1 < C_2$. For $m = 0$ we will see below that the answer is in the negative For general m the subject is under investigation.

In [7] (Theorem 3) the present author gave an extremely simple and very much briefer proof of Theorem B. Using essentially the same simple methods he proved the following improvement on Theorem A.

[5] More about error correcting codes in, for example, [6]

THEOREM 1 of [7]: *For any n there exists a code of length*

$$2^{nC_1-K_1n^{1/2}},$$

where K_1 is a positive constant[6] *which does not depend on n.*

We next concern ourselves with the important and interesting question of an upper bound for the length of an (error correcting) code. For the codes considered by Shannon the latter stated[7] ([1]) that there cannot exist a code of length greater than

$$2^{n(C_2+\epsilon)}.$$

Shannon gave a proof to which all others in the literature refer. Khintchine [4] pointed out that neither the argument of [1] nor any of the arguments to be found in the literature constitute a proof or even the outline of a proof; he also pointed out the desirability of proving the result and mentioned some of the difficulties.

In [7] (Theorem 2) the author proved the following theorem: *When $m = 0$ there is a positive constant*[8] *K_2 such that there cannot exist an (error correcting) code of length greater than*

$$2^{nC_0+K_2n^{1/2}}.$$

An immediate consequence of this theorem is that, when $m = 0$, $C_0 = C_1 = C_2$. Hence, when $m = 0$, Theorem B adds nothing to Theorem A, and both are weaker than Theorem 1.

Before passing to the case $m > 0$ we complete the above discussion by defining the functional φ. Let

$$X = (X_1, \cdots, X_n)$$

and define

$$Y(X) = (Y_1, \cdots, Y_{n-m}) = Y \text{ (say)}.$$

(More detailed definitions in [7]; Y is essentially the chance sequence received when the chance sequence X is sent.) Define the symbol

$$P\{Y = y \mid X = x\} \; (P\{X = x \mid Y = y\})$$

as the conditional probability that $Y = y$, given $X = x$ (that $X = x$, given $Y = y$). We define the following functions of the chance variables X and Y: When $X = x$ and $Y = y$, then

function	*equals*
$P\{X\}$	$P\{X = x\}$
$P\{Y\}$	$P\{Y = y\}$
$P\{X \mid Y\}$	$P\{X = x \mid Y = y\}$
$P\{Y \mid X\}$	$P\{Y = y \mid X = x\}$

[6] K_1 depends on the channel probability function.

[7] There is some ambiguity about the theorem actually stated.

[8] K_2 depends on the channel probability function.

Let E denote the expected value operator. It is proved that the following limits all exist:

$$\lim -\frac{1}{n} E[\log P\{X\}] = D_1$$

$$\lim -\frac{1}{n} E[\log P\{Y\}] = D_2$$

$$\lim -\frac{1}{n} E[\log P\{X \mid Y\}] = D_3$$

$$\lim -\frac{1}{n} E[\log P\{Y \mid X\}] = D_4$$

Also it is true and easy to prove that

$$D_1 + D_4 = D_2 + D_3$$

Then

$$\varphi = D_1 - D_3 = D_2 - D_4$$

We now turn to the general case $m \geqq 0$. For this case Theorem 4 of [8] gives a general upper bound for the length of an (error correcting) code. When $m = 0$ Theorem 4 specializes to Theorem 2. Whether Theorem 4 gives the "best" upper bound (as Theorem 2 does for $m = 0$) is still under investigation. Unfortunately, to state Theorem 4 one needs a page of preliminary definitions and then the theorem is stated in terms which require the reader to be familiar with the theory of Markov chains. (However, the application of the theorem as described in the discussion of [8] which follows its proof is little more difficult than that of Theorem 2). For these reasons it seems best to refer the interested reader to [8].

Postscript added in December, 1957.

Since this paper was submitted for publication the author has obtained the following result: A number J is defined by means of certain algebraic and analytic operations on the channel probability function which we shall not describe here. For any positive ϵ and n sufficiently large, there exists a code of length $2^{n(J-\epsilon)}$, and there cannot exist a code of length greater than $2^{n(J+\epsilon)}$. This result can be approximately described by saying that 2^{nJ} is the maximum achievable code length.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. Jour.* Vol. 27 (1948), pp. 379–423, 623-656.

[2] A. Feinstein, "A new basic theorem of information theory," *Trans. Inst. Radio Eng.*, Professional Group on Information Theory (1954), pp. 2-22.

[3] A. Khintchine, "The concept of entropy in the theory of probability," *Uspekhi Mat. Nauk* VIII, 3(55), (1953), 3-20.

[4] A. Khintchine, "On the fundamental theorems of the theory of information," Uspyekhi Mat. Nauk XI, 1(67), (1956), 17–75.
[5] B. McMillan, "The basic theorems of information theory," Ann. Math. Stat. 24, No. 2 (1953), 196–219.
[6] E. N. Gilbert, "A comparison of signaling alphabets," Bell System Tech. Journal, 31 (1952), pp. 504–522.
[7] J. Wolfowitz, "The coding of messages subject to chance errors," to appear in Illinois Journal of Mathematics.
[8] J. Wolfowitz, "An upper bound on the rate of transmission of messages," to appear in Illinois Journal of Mathematics.

SOME PROBLEMS CONNECTED WITH STATISTICAL INFERENCE

By D. R. Cox

Birkbeck College, University of London[1]

1. Introduction. This paper is based on an invited address given to a joint meeting of the Institute of Mathematical Statistics and the Biometric Society at Princeton, N. J., 20th April, 1956. It consists of some general comments, few of them new, about statistical inference.

Since the address was given publications by Fisher [11], [12], [13], have produced a spirited discussion [7], [21], [24], [31] on the general nature of statistical methods. I have not attempted to revise the paper so as to comment point by point on the specific issues raised in this controversy, although I have, of course, checked that the literature of the controversy does not lead me to change the opinions expressed in the final form of the paper. Parts of the paper are controversial; these are not put forward in any dogmatic spirit.

2. Inferences and decisions. A statistical inference will be defined for the purposes of the present paper to be a statement about statistical populations made from given observations with measured uncertainty. An inference in general is an uncertain conclusion. Two things mark out statistical inferences. First, the information on which they are based is statistical, i.e. consists of observations subject to random fluctuations. Secondly, we explicitly recognise that our conclusion is uncertain, and attempt to measure, as objectively as possible, the uncertainty involved. Fisher uses the expression 'the rigorous measurement of uncertainty'.

A statistical inference carries us from observations to conclusions about the populations sampled. A scientific inference in the broader sense is usually concerned with arguing from descriptive facts about populations to some deeper understanding of the system under investigation. Of course, the more the statistical inference helps us with this latter process, the better. For example, consider an experiment on the effect of various treatments on the macroscopic properties of a polymer. The statistical inference is concerned with what can be inferred from the experimental results about the true treatment effects. The scientific inference might concern the implications of these effects for the molecular structure of the polymer; the statistical uncertainty is only a part, sometimes small, of the uncertainty of the final inference.

Statistical inferences, in the sense meant here, involve the data, a specification of the set of possible populations sampled and a question concerning the true populations. No consideration of losses is usually involved directly in the inference, although these may affect the question asked. If the population sampled

Received October 7, 1957; revised February 10, 1958.

[1] Work done at the Department of Biostatistics, School of Public Health, University of North Carolina.

has itself been selected by a random procedure with known prior probabilities, it seems to be generally agreed that inference should be made using Bayes's theorem. Otherwise, prior information concerning the parameter of direct interest[2] will not be involved in a statistical inference. The place of prior information is discussed some more when we come to talk about decisions, but the general point is that prior information that is not statistical cannot be included without abandoning the frequency theory of probability, and information that is derived from other statistical data can be handled by methods for the combination of data.

The theory of statistical decision deals with the action to take on the basis of statistical information. Decisions are based on not only the considerations listed for inferences, but also on an assessment of the losses resulting from wrong decisions, and on prior information, as well as, of course, on a specification of the set of possible decisions. Current theories of decision do not give a direct measure of the uncertainty involved in making the decision; as explained above, a statistical inference is regarded here as having an explicitly measured uncertainty, and this is to be thought of as an essential distinction between statistical decisions and statistical inferences.

Thus, significance tests and confidence intervals, if looked at in the way explained below, are inference procedures. Discriminant analysis, considered as a method for classifying individuals into one of two groups, is a decision procedure; considered as a tool for assigning a score to an individual to say how reasonable it is that the individual comes from one group rather than the other, it is an inference procedure. Strict point estimation represents a decision; estimation by point estimate and standard error is a condensed and approximate form of interval estimation and is an inference procedure. Estimation by a posterior distribution derived from an agreed prior distribution is an inference procedure. A test of a hypothesis, considered in the literal Neyman-Pearson sense as a rule for taking one of two decisions concerning a statistical hypothesis, is a decision procedure, in which prior knowledge and losses enter implicitly. The reader may find it helpful to consider the extent to which the specification, implicitly or explicitly, of losses and prior knowledge is essential for solution of the problems just listed as ones of decision.

For example, consider the analysis of an experiment to compare two industrial processes, A and B. The statistical inference might be that, under certain assumptions about the populations, process A gives a yield higher than that of process B, the difference being statistically significant past the 1/1000 level, 90, 95 and 99 per cent confidence intervals for the amount of the true difference being such and such. The decision might be that having regard to the differences in yield of practical importance, and our prior knowledge, we will consider that the experiment has established, under the conditions examined, that process A has a higher yield than B and will take future action accordingly.

[2] *i.e.* relevant information about the parameter of interest, other than that contained in the data and in the specification of the set of possible parameter values.

An inference without a prior distribution can be considered as answering the question: 'What do these data entitle us to say about a particular aspect of the populations that interest us?' It is, however, irrational to take action, scientific or technological, without considering both all available relevant information, including for example the prior reasonableness of different explanations of a set of data, and also the consequences of doing the wrong thing. Why then, do we bother with inferences which go, as it were, only part of the way towards the final decision?

Even in problems where a clear-cut decision is the main object, it very often happens that the assessment of losses and prior information is subjective, so that it will help to get clear first the relatively objective matter of what the data say, before embarking on the more controversial issues In particular, it may happen either that the data are little aid in deciding the point at issue, or that the data suggest one conclusion so strongly that the only people in doubt about what to do are those with prior beliefs, or opinions about losses, heavily biased in one direction. In some fields, too, it may be argued that one of the main calls for probabilistic statistical methods arises from the need to have agreed rules for assessing strength of evidence.

A full discussion of this distinction between inferences and decisions will not be attempted here. Three more points are, however, worth making briefly. First, some people have suggested that what is here called inference should be considered as 'summarization of data'. This choice of words seems not to recognise that an essential element is the uncertainty involved in passing from the observations to the underlying populations[1] Secondly, the distinction drawn here is between the applied problem of inference and the applied problem of decision-making; it is possible that a satisfactory set of techniques for inference could be constructed from a mathematical structure very similar to that used in decision theory.

Finally, it might be argued that in making an inference we are 'deciding' to make a statement of a certain type about the populations and that therefore, provided that the word decision is not interpreted too narrowly, the study of statistical decisions embraces that of inference. The point here is that one of the main general problems of statistical inference consists in deciding what types of statement can usefully be made and exactly what they mean. In statistical decision theory, on the other hand, the possible decisions are considered as already specified.

3. The sample space. Statistical methods work by referring the observations S to a sample space Σ of observations that might have been obtained. Over Σ one or more probability measures are defined and calculations in these probability distributions give our significance limits, confidence intervals, etc. Σ is usually taken to be the set of all possible samples having the same size and structure as the observations.

[1] A referee has suggested the term 'summarization of evidence,' which seems a good one

Fisher (see, for example, [11]) and Barnard [4] have pointed out that Σ may have no direct counterpart in indefinite repetition of the experiment. For example, if the experiment were repeated, it may be that the sample size would change. Therefore what happens when the experiment is repeated is not sufficient to determine Σ, and the correct choice of Σ may need careful consideration.

As a comment on this point, it may be helpful to see an example where the sample size is fixed, where a definite space Σ is determined by repetition of the experiment and yet where probability calculations over Σ do not seem relevant to statistical inference.

Suppose that we are interested in the mean θ of a normal population and that, by an objective randomization device, we draw either (i) with probability $\frac{1}{2}$, one observation, x, from a normal population of mean θ and variance σ_1^2 or (ii) with probability $\frac{1}{2}$, one observation x, from a normal population of mean θ and variance σ_2^2, where σ_1^2, σ_2^2 are known, $\sigma_1^2 \gg \sigma_2^2$ and where we know in any particular instance which population has been sampled.

More realistic examples can be given, for instance in terms of regression problems in which the frequency distribution of the independent variable is known. However, the present example illustrates the point at issue in the simplest terms. (A similar example has been discussed from a rather different point of view in [6], [29]).

The sample space formed by indefinite repetition of the experiment is clearly defined and consists of two real lines Σ_1, Σ_2, each having probability $\frac{1}{2}$, and conditionally on Σ_i there is a normal distribution of mean θ and variance σ_i^2.

Now suppose that we ask, accepting for the moment the conventional formulation, for a test of the null hypothesis $\theta = 0$, with size say 0.05, and with maximum power against the alternative θ', where $\theta' \simeq \sigma_1 \gg \sigma_2$.

Consider two tests. First, there is what we may call the conditional test, in which calculations of power and size are made conditionally within the particular distribution that is known to have been sampled. This leads to the critical regions $x > 1.64\ \sigma_1$ or $x > 1.64\ \sigma_2$, depending on which distribution has been sampled.

This is not, however, the most powerful procedure over the whole sample space. An application of the Neyman-Pearson lemma shows that the best test depends slightly on θ', σ_1, σ_2, but is very nearly of the following form. Take as the critical region

$$x > 1.28\sigma_1, \qquad \text{if the first population has been sampled;}$$

$$x > 5\sigma_2, \qquad \text{if the second population has been sampled.}$$

Qualitatively, we can achieve almost complete discrimination between $\theta = 0$ and $\theta = \theta'$ when our observation is from Σ_2, and therefore we can allow the error rate to rise to very nearly 10% under Σ_1. It is intuitively clear, and can easily be verified by calculation, that this increases the power, in the region of interest, as compared with the conditional test.

Now if the object of the analysis is to make statements by a rule with certain

specified long-run properties, the unconditional test just given is in order, although it may be doubted whether the specification of desired properties is in this case very sensible. If, however, our object is to say 'what we can learn from the data that we have', the unconditional test is surely no good. Suppose that we know we have an observation from Σ_1. The unconditional test says that we can assign this a higher level of significance than we ordinarily do, because if we were to repeat the experiment, we might sample some quite different distribution. But this fact seems irrelevant to the interpretation of an observation which we know came from a distribution with variance σ_1^2. That is, our calculations of power, etc. should be made conditionally within the distribution known to have have been sampled, i.e. if we are using tests of the conventional type, the conditional test should be chosen.

To sum up, if we are to use statistical inferences of the conventional type, the sample space Σ must not be determined solely by considerations of power, or by what would happen if the experiment were repeated indefinitely. If difficulties of the sort just explained are to be avoided, Σ should be taken to consist, so far as is possible, of observations similar to the observed set S, in all respects which do not give a basis for discrimination between the possible values of the unknown parameter θ of interest. Thus, in the example, information as to whether it was Σ_1 or Σ_2 that we sampled tells us nothing about θ, and hence we make our inference conditionally on Σ_1 or Σ_2.

Fisher has formalized this notion in his concept of ancillary statistics [10], [23], [27]. His definitions deal with the situation without nuisance parameters and before outlining an extension that attempts to cope with nuisance parameters, it is convenient to state a slight modification of the original definitions. Let **m** be a minimal set of sufficient statistics[4] for the unknown parameter of interest, **θ**, and suppose that **m** can be written (**t**, **a**), where the distribution of **a** is independent of **θ**, and that no further components can be extracted from **t** and incorporated in **a**. That is, we divide, if possible, the space of **m** into sets each similar to the sample space, and take the finest such division, assumed here to be unique subject to regularity conditions. Then **a** is called an ancillary statistic and we agree to make inferences conditionally on the observed **a**.

Examples. (i) In the example of section 3, a minimal set consists of the observation, x, and an indicator variable to show which population has been sampled. The latter satisfies the conditions for being an ancillary statistic. Provided that the possible values of the mean θ include an interval, there is no set of x values with the same probability for all θ.

(ii) Under the ordinary assumptions of normal linear regression theory, plus the assumption that the independent variable has any known distribution (without unknown parameters), the values of the independent variable form an ancillary statistic.

(iii) The following example is derived from one put forward by a referee.

[4] The terms used by Fisher are that a minimal set of sufficient statistics with more components than there are parameters is called *exhaustive* and a minimal set with the same number of components as there are parameters is called *sufficient*.

Let x be a single observation with density $1 + 2\theta x$, $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$, $-1 \leqq \theta \leqq 1$. Then we can write $x = [\text{sgn } x, |x|]$ and $|x|$ has the same density for all θ. Hence we argue conditionally on the observed value of $|x|$. For example in testing $\theta = 0$ against $\theta > 0$, the possible P values (see section 5) are 1 and $\frac{1}{2}$. This may seem a curious result but is, I think, reasonable if one regards a significance test as concerned with the extent to which the data are consistent with the null hypothesis.

Suppose now that there are nuisance parameters $\boldsymbol{\phi}$. Let $\mathbf{m}$ be a minimal set of sufficient statistics for estimating $(\boldsymbol{\theta}, \boldsymbol{\phi})$ and suppose that $\mathbf{m}$ can be partitioned into $[\mathbf{t}, \mathbf{s}, \mathbf{a}]$ in such a way that

(i) functions of $\mathbf{t}$ and $\boldsymbol{\theta}$, so-called pivotal quantities, exist with a distribution conditionally on $\mathbf{a}$ that is independent of $\boldsymbol{\phi}$. If any component of $\mathbf{s}$ is added to $\mathbf{t}$ or $\mathbf{a}$, this independence from $\boldsymbol{\phi}$ no longer holds. Further, no components can be extracted from $\mathbf{t}$ and incorporated in $\mathbf{a}$;

(ii) the values of $\mathbf{a}$ and $\mathbf{s}$ give no direct information about $\boldsymbol{\theta}$ in the sense to be defined below. Then we agree to make inferences about $\boldsymbol{\theta}$ from the conditional distribution of (i).

We need then to define what is meant by saying that a quantity $\mathbf{y}$ gives no direct information about $\boldsymbol{\theta}$, when nuisance parameters $\boldsymbol{\phi}$ are present. One condition that might be considered is that the density $p(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\phi})$ should be independent of $\boldsymbol{\theta}$. This seems too strong, as does also the requirement that for every different pair $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and for every $\mathbf{y}$, $p(\mathbf{y}; \boldsymbol{\theta}_1, \boldsymbol{\phi}) / p(\mathbf{y}; \boldsymbol{\theta}_2, \boldsymbol{\phi})$ should run through all positive real values as $\boldsymbol{\phi}$ varies. An appropriate condition seems to be that given admissible values $\mathbf{y}$, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\phi}$, there exist admissible $\boldsymbol{\theta}$, $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, such that

$$\frac{p(\mathbf{y}; \boldsymbol{\theta}_1, \boldsymbol{\phi})}{p(\mathbf{y}; \boldsymbol{\theta}_2, \boldsymbol{\phi})} = \frac{p(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\phi}_1)}{p(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\phi}_2)}. \tag{1}$$

The import of the condition is that any contemplated distinction between two values of $\boldsymbol{\theta}$ might just as well be regarded as a distinction between two values of $\boldsymbol{\phi}$.

For example, suppose that x is a single observation from a normal distribution of unknown mean ϕ and variance θ. Then x gives no direct information about θ in the sense of (1), provided that ϕ is completely unknown. Another example is normal regression theory with the independent variable having an arbitrary unknown distribution, not involving the regression parameters of interest [10]. Here $\mathbf{a}$ is the set of values of the independent variable and s is the sum squares about the regression line, assuming that the residual variance about the regression line, ϕ, is a nuisance parameter.

For a third example, let r_1, r_2 be randomly drawn from Poisson distributions of means μ_1, μ_2 and let $\mu_2/\mu_1 = \theta$ be the parameter of interest; that is write the means as ϕ, $\phi\theta$, where ϕ is a nuisance parameter. The likelihood of r_1, r_2 can be written

$$\frac{e^{-\phi(1+\theta)}[\phi(1+\theta)]^a}{a!} \times \frac{a!}{t!(a-t)!}\left(\frac{1}{1+\theta}\right)^t\left(\frac{\theta}{1+\theta}\right)^{a-t},$$

where $t = r_1$, $a = r_1 + r_2$ and with s null. The equation (1) is satisfied, telling us that a gives us no direct information about θ. Therefore significance and confidence calculations are to be made conditionally on the observed value of a, as is the conventional procedure [25].

To apply the definitions we have to regard our observations as generated by a random process; the idea of ancillary statistics simply tells us how to cut down the sample space to those points relevant to the interpretation of the observations we have.

In the problems without nuisance parameters, it is known that methods of inference [5], that use only observed values of likelihood ratios, and not tail areas, avoid the difficulties discussed above, since the likelihood ratio is the same whether we argue conditionally or not. Lindley, using concepts from [18], has recently shown that for a broad class of problems with nuisance parameters, the conditional methods are optimum in the Neyman-Pearson sense.

Another important problem connected with the choice of the sample space, not discussed here, concerns the possibility and desirability of making inferences within finite sample spaces obtained by permuting the observations; see, for example, [16].

4. Interval estimation. Much controversy has centred on the distinction between fiducial and confidence estimation. Here follow five remarks, not about the mathematics, but about the general aims of the two methods

(i) The fiducial approach leads to a distribution for the unknown parameter, whereas the method of confidence intervals, as usually formulated, gives only one interval at some preselected level of probability. This seems at first sight a distinct point in favour of the fiducial method. For when we write down the confidence interval $(\bar{x} - 1.96\,\sigma/\sqrt{n}, \bar{x} + 1.96\,\sigma/\sqrt{n})$ for a completely unknown normal mean, there is certainly a sense in which the unknown mean θ is likely to lie near the centre of the interval, and rather unlikely to lie near the ends and in which, in this case, even if θ does lie outside the interval, it is probably not far outside. The usual theory of confidence intervals gives no direct expression of these facts.

Yet this seems to a large extent a matter of presentation; in the common simple cases, where the upper α limit for θ is monotone in α, there seems no reason why we should not work with confidence distributions for the unknown parameter. These can either be defined directly, or can be introduced in terms of the set of all confidence intervals at different levels of probability. Statements made on the basis of this distribution, provided we are careful about their form, have a direct frequency interpretation. In applications it will often be enough to specify the confidence distribution, by for example a pair of intervals, and this corresponds to the common practice of quoting say both the 95 per cent and the 99 per cent confidence intervals.

It is not clear what can be done in those complex cases [8], [26], where say the upper 5 per cent limit for θ is larger than the upper 1 per cent limit, or indeed whether confidence interval estimation is at all satisfactory in such cases.

Within the class of distributions with monotone likelihood ratio [15], such difficulties will, however, be avoided.

If we consider that the object of interval estimation is to give a rule for making on the basis of each set of data, a statement about the unknown parameter, a certain preassigned proportion of the statements to be correct in the long run, consideration of the confidence distribution may seem unnecessary and possibly invalid. The attitude taken here is that the object is to attach, on the basis of data S, a measure of uncertainty to different possible values of θ, showing what can be inferred about θ from the data. The frequency interpretation of the confidence intervals is the way by which the measure of uncertainty is given a concrete interpretation, rather than the direct object of the inference. From this point of view it is difficult to see an objection to the consideration of many confidence statements simultaneously.

If the whole set of intervals is regarded as the fundamental concept, and if we are interested both in upper and in lower limits for θ, we may conveniently specify the set by giving say the upper and lower $2\frac{1}{2}$% points, etc., it being a useful convention, and no more, that the 95% interval so obtained should have equal probabilities associated with each tail. The elaborate discussion that is sometimes necessary in the conventional theory to decide which particular combination of upper and lower tail areas is best to get a 95% interval seems, from this point of view, irrelevant.

(ii) It is sometimes claimed as an advantage of fiducial estimation that it is restricted to methods that use 'all the information in the data', while confidence estimation includes any method giving the requisite frequency interpretation. This claim is lent some support by those accounts of confidence interval theory which use the words 'valid' or 'exact' for a method of calculating intervals that has, under a given mathematical set-up, an exact frequency interpretation, no matter how inadequate the intervals may be in telling us what can be learnt from the data.

However, good accounts of the theory of confidence intervals stress equally the need to cover the true value with the required probability and the requirement of having the intervals as narrow as possible in a suitable sense [21]. Very special importance, therefore, attaches to intervals based on exhaustive estimates. It is true that there are differences between the approaches in that the fiducial method takes the use of exhaustive estimates as a primary requirement, whereas in the theory of confidence intervals the use of exhaustive estimates is deduced from some other condition. This does not seem however to amount to a major difference between the methods.

(iii) The uniqueness of inferences obtained by the fiducial method has received much discussion recently, [9], [20], [28]. Uniqueness is important because, once the mathematical form of the populations is sufficiently well specified, it should be possible to give a single answer of a given type to the question 'what do the data tell us about θ?'.

The present position is that several cases are known where the fiducial method leads to non-unique answers, although it is, of course, entirely possible that a way will be found of formulating fiducial calculations to make them unique. A comparison with confidence intervals is difficult here, because in many of the multi-parameter problems, the single parameters for which confidence estimation is known to be possible at all are very limited. No cases of non-unique optimum confidence intervals seem to have been published.

(iv) If sufficient estimation, in Fisher's sense, is possible for a group of parameters, fiducial inference will usually be possible about any one of them or any combination of them, since the joint fiducial distribution of all the parameters can be found and the unwanted parameters integrated out. [illegible] confidence estimation is in general possible only for restricted combinations of parameters. An example is the Behrens-Fisher problem, where exact fiducial inference is possible. The situation about confidence estimation in this case is far from clear, but may be that the asymptotic expansion proposed by Welch [30], while giving a close approximation to an 'exact' system of confidence intervals, has frequency properties depending slightly on the nuisance parameter. Nothing seems to be known about possible optimum properties in the [illegible] Pearson sense. In the language of testing hypotheses, Welch's procedure [illegible] look for a region of constant size α, independently of the nuisance parameter. It is conceivable that greater power against some alternatives [illegible] having a size only bounded by α; indeed, this is made plausible by [illegible]

(v) The final consideration concerns the question of frequency [illegible] Fisher has repeatedly stated that the *immediate* object of fiducial [illegible] not the making of statements that will be correct with given frequency [illegible] long run. One may readily accept this in that one really wants to [illegible] uncertainty corresponding to different ranges of values for [illegible] quite conceivable that one could construct a satisfactory measure of uncertainty [illegible] that has not a direct frequency interpretation. Yet one must surely [illegible] some pretty clear-cut practical meaning to the measure of uncertainty [illegible] this fiducial probability has never been shown to have, except in [illegible] where it is equivalent to confidence interval estimation. [illegible] recent unpublished work on fiducial probability and its frequency [illegible] may be mentioned here.

A different justification of fiducial distributions [illegible] advanced is to derive them from Bayes's theorem, using a [illegible] form of prior distribution. To remain within the framework of the frequency theory of probability, it would then be necessary to distinguish between proper frequency distributions and hypothetical ones. The physical interpretation of the measure of uncertainty of statements about θ is that if θ had such and such a prior frequency distribution, then the posterior frequency distribution of θ would be such and such. This all amounts to a reinterpretation of Jeffreys's theory. An important advantage of this approach is that it ensures indep[illegible]

the sampling rule (see [2]) and from the difficulties of section 3. On the other hand it seems a clumsy way of dealing with simple one-parameter problems, especially when the choice of prior distribution is difficult.

If the above considerations are accepted, it seems reasonable to base interval estimation on a slightly revised form of the theory of confidence intervals.

Estimation by confidence or fiducial distribution may be contrasted with the proposal [5], [13] to plot the likelihood of the unknown parameter θ in the light of the data, standardized by the maximum likelihood over θ. Advantages of the latter method are mathematical simplicity and independence from the sampling rule. Disadvantages are that it is not clear how to deal with nuisance parameters, that it is not clear that division by the maximum value of the likelihood makes values in different situations genuinely comparable, and that there is some difficulty in giving practical interpretation to the ratios so obtained. It might be argued that this last difficulty arises solely from lack of familiarity with the method.

5. Significance tests. Suppose now that we have a null hypothesis H_0 concerning the population or populations from which the data S were drawn and that we enquire 'what do the data tell us concerning the possible truth or falsity of H_0?' Adopt as a measure of consistency with the null hypothesis

$$\text{prob}\left\{\begin{matrix}\text{data showing as much or more}\\ \text{evidence against } H_0 \text{ as } S\end{matrix}\,\middle|\, H_0\right\}. \tag{2}$$

That is, we calculate, at least approximately, the actual level of significance attained by the data under analysis and use this as a measure of conformity with the null hypothesis. The value obtained in this way is often, particularly in the biological literature, called the P-value. Significance tests are often used in practice like this, although many formal accounts of the theory of tests suggest, implicitly or explicitly, quite a different procedure. Namely, we should, after considering the consequences of wrongly accepting and rejecting the null hypothesis, and the prior knowledge about the situation, fix a significance level in advance of the data. This is then used to form a rigid dividing line between samples for which we accept the null hypothesis and those for which we reject the null hypothesis. A decision-type of this sort is clearly something quite different from the application just contemplated.

Two aspects of significance tests will be discussed briefly here. First there is the question of when significance tests are useful and secondly there is the justification of (2) as a measure of conformity.

We shall for simplicity, consider situations in which the possible populations correspond to values of a continuously varying parameter θ, the null hypothesis being say $\theta = \theta_0$. There may be nuisance parameters.

A practical distinction can be made between cases in which the null value θ_0 is considered because it divides the parameter range into qualitatively different

sections and those cases in which it is thought that there is a reasonable prospect that the null value is very nearly the true one. For example, in the comparison of two alternative industrial processes we might quite often have no particular expectation that the treatment difference is small. In such cases the significance test is concerned with whether we can, *from the data under analysis*, claim the existence of a difference in the same direction as that observed. Or, to look at the matter slightly differently, the significance level tells us at what levels the confidence intervals for the true difference include only values with the same sign as the sample difference. This idea that the significance level is concerned with the possibility that the true effect may be in the opposite direction from that observed, occurs in a different way in [17].

The answer to the significance test is rarely the only thing we should consider: whether or not significance is attained at an interesting level (say at the 10% level or better), some consideration should be given to whether differences that may exist are of practical importance, i.e. estimation should be considered as well as significance testing. A likely exception to this is in the analysis of rather limited amounts of data, where it can be taken for granted that differences of practical importance are consistent with the data. The point of the statistical analysis is in such cases to see whether the direction of any effects has been reasonably well established, i.e. whether a qualitative conclusion about the effects has been demonstrated.

The problem dealt with by a significance test, as just considered, is different from that of deciding which of two treatments is to be recommended for future use or further investigation. This cannot be tackled without consideration of the differences of practical importance, the losses consequent on wrong decisions and the prior knowledge. Depending on these and on sample size, the level of P for practical action may vary widely.

The second type of application of significance tests is to situations where there is a definite possibility that the null hypothesis is nearly true. (Exact truth of a null hypothesis is very unlikely except in a genuine uniformity trial). A full analysis of such a situation would involve consideration of what departure from the null hypothesis is considered of practical importance. However, it is often convenient to test the null hypothesis directly; if significant departure from it is obtained, consideration must then be given to whether the departure is of practical importance. Of course, in any case we will probably wish to examine the problem as one of estimation as well as of significance testing, asking for example, for the maximum true difference consistent with the data.

Consider now the choice of (2) as the quantity to measure significance. To use the definition, we need to order the points of the sample space in terms of the evidence they provide against the null hypothesis.

The most satisfactory way is the introduction, as in the usual development of the Neyman-Pearson theory, of the requirement of maximum sensitivity in the detection of certain types of departure from the null hypothesis. That is, we wish, in the simplest case, to maximise, if possible for all fixed ϵ,

$$\text{prob}_\theta(\text{attaining significance at the } \epsilon \text{ level}),$$

where θ represents a set-up which we desire to distinguish from the null hypothesis. That is we choose the procedure that makes the random variable (2) as stochastically small as possible when the alternative hypotheses are true. This leads in simple cases, to a unique specification of the significance probability (2).

In the simple case when there is a single alternative hypothesis, it seems at least of theoretical interest to distinguish between the problem of discrimination and that of significance testing. In discrimination, the two populations are on an equal footing and there are strong arguments for considering that only the observed value of the likelihood ratio is relevant. The question asked is 'which of these populations do the observations come from?' In significance testing the question is 'are the data consistent with having come from H_0?' The alternative hypothesis serves merely to mark out the sample points giving evidence against H_0.

The next question to consider is why we sum over a whole set of sample points rather than work in terms only of the observed point. This has been much discussed. The advantage of (2) is that it has a clear-cut physical interpretation in terms of the formal scheme of acceptance and rejection contemplated in the Neyman-Pearson theory. To obtain a measure depending only on the observed sample point, one way is to take the likelihood ratio, for the observed point, of the null hypothesis versus some conventionally chosen alternative (see [5]), and while a practical meaning can be given to this, it has less direct appeal. But consider a test of the following discrete null hypotheses:

Sample value	prob. under H_0	prob. under H_0'
0	0.80	0.75
1	0.15	0.15
2	0.05	0.05
3	0.00	0.01
4	0.00	0.01

and suppose that the alternatives are the same in both cases and are such that the probabilities (2) should be calculated by summing the probabilities of values as great or greater than that observed. Suppose further that the observation 2 is obtained; under H_0 the significance level is 0.05, while under H_0' it is 0.10. Yet it is difficult to see why we should say that our observation is more consistent with H_0 than with H_0'; this point has often been made before [4], [16]. On the other hand, if we are really interested in the confidence interval type of problem, i.e. in covering ourselves against the possibility that the 'effect' is in the direction opposite to that observed, the use of the tail area seems more reasonable. As noted in section 3 the use of likelihood ratios rather than summed probabilities avoid difficulties connected with the choice of the sample space, Σ. We are faced with a conflict between the mathematical and logical advantages of the likelihood ratio, and the desire to calculate quantities with a clear practical meaning in terms of what happens when they are calculated.

In general the role that tail areas ought to play in statistical inference is far from clear and further discussion is very desirable. The reader may refer to [1] and [19].

In this and the preceding section the problems of interval estimation and significance testing have been considered. There is not space to give a parallel discussion of the other types of statistical procedure.

6. The role of the assumptions. The most important general matter connected with inference not discussed so far, concerns the role of the assumptions made in calculating significance, etc. Only a very brief account of this matter will be given here.

Assumptions that we make, such as those concerning the form of the populations sampled, are always untrue, in the sense that, for example, enough observations from a population would surely show some systematic departure from say the normal form. There are two devices available for mitigating this difficulty, namely

(i) the idea of nuisance parameters, i.e. of inserting sufficient unknown parameters into the functional form of the population, so that a better approximation to the true population can be attained;

(ii) the idea of robustness (or stability), i.e. that we may be able to show that the answer to the significance test or estimation procedure would have been essentially unchanged had we started from a somewhat different population form. Or, to put it more directly, we may attempt to say how far the population would have to depart from the assumed form, to change the final conclusions seriously. This leaves us with a statement that has to be interpreted qualitatively in the light of prior information about distributional shape, plus the information, if any, to be gained from the sample itself. This procedure is frequently used in practical work, although rarely made explicit.

In inference for a single population mean, examples of (i) are, in order of complexity, to assume

(a) a normal population of unknown dispersion;
(b) a population given by the first two terms of an Edgeworth expansion;
(c) in the limit, either an arbitrary population, or an arbitrary continuous population (leading to a distribution-free procedure).

The last procedure has obvious attractions, but it should be noted that it is not possible to give a firm basis for choice between numerous alternative methods, without bringing in strong assumptions about the power properties required, and also that it often happens that no reasonable distribution-free method exists for the problem of interest. Thus if we are concerned with the mean of a population of unknown shape and dispersion, no distribution-free method is available [3]; when the property measured is extensive, the mean is often the uniquely appropriate parameter.

A rather artificial example of method (ii) is that if we were given a single observation from a normal population and asked to assess the significance of the difference from zero, we could plot the level attained against the population

standard deviation σ. Then we could interpret this qualitatively in the light of whatever prior information about σ was available. A less artificial example concerns the comparison of two sample variances. The ratio might be shown to be highly significant by the usual F test and a rough calculation made to show that provided that neither β_2 exceeded β_2^0, significance at least say at the 1 per cent level would still occur.

In practical situations we usually employ a mixture of (i) and (ii) depending on

(a) the extent to which our prior knowledge limits the population form in respects other than those of direct interest;
(b) the amount of information in the data about the population characteristic that may be used as a nuisance parameter;
(c) the extent to which the final conclusion is sensitive to the particular population characteristic of interest.

Thus, in (a) if we have a good idea of the population form, we are probably not much interested in the fact that a distribution-free method has certain desirable properties for distributions quite unlike that we expect to encounter. To comment on (b), we would probably not wish to studentize with respect to a minor population characteristic about which hardly any information was contained in the sample, e.g. an estimate of variance with one or two degrees of freedom. In small sample problems there is frequently little information about population shape contained in the data. Finally, there is consideration (c). If the final conclusion is very stable under changes of distribution form, it is usually convenient to take the most appropriate simple theoretical form as a basis for the analysis and to use method (ii).

Now it is very probable that in many instances investigation would show that the same answer would, for practical purposes, result from the alternative types of method we have been discussing. But suppose that in a particular instance there is disagreement, e.g. that the result of applying a t test differs materially from that of applying some distribution-free procedure. What should we do?

It can be argued that, even if we have no good reason for expecting a normal population, we should not be willing to accept the distribution-free answer unconditionally. A serious difference between the results of the two tests would indicate that the conclusion we draw about the population mean depends on the population shape in an important way, e.g. depends on the attitude we take to certain outlying observations in the sample. It seems more satisfactory for a full discussion of the data, to state this and to assemble whatever evidence is available about distributional form, rather than simply to use the distribution-free approach. Distribution-free methods are, however, often very useful in small sample situations where little is known about population form and where elaborate treatment of the results would be out of place.

An interesting discussion of the role of assumptions in decision theory is given in [14].

I am much indebted to the two referees for detailed and constructive criticism of the paper.

REFERENCES

[1] F. J. ANSCOMBE, "Contribution to the discussion of a paper by F N. David and N. L Johnson" *J.R. Stat. Soc.*, B, Vol 18, (1956), pp. 24–27.

[2] F. J. ANSCOMBE, "Dependence of the fiducial argument on the sampling rule", *Biometrika*, Vol. 44 (1957), pp. 464–469.

[3] R. R. BAHADUR AND L. J. SAVAGE, "The non-existence of certain statistical procedures in non-parametric problems", *Ann. Math. Stat.*, Vol. 27 (1956), pp 1115–1122.

[4] G. A. BARNARD, "The meaning of a significance level", *Biometrika*, Vol 34 (1947), pp. 179–182.

[5] G A. BARNARD, "Statistical inference", *J R. Stat. Soc., Suppl*, Vol 11 (1949), pp 115–139.

[6] M. S BARTLETT, "A note on the interpretation of quasi-sufficiency", *Biometrika*, Vol. 31 (1939), pp 391–392

[7] M. S. BARTLETT, "Comment on Sir Ronald Fisher's paper", *J.R Stat Soc.*, B, Vol 18 (1956), pp. 295–296.

[8] H. CHERNOFF, "A property of some type of A regions", *Ann Math Stat*, Vol 22 (1951), pp 472–474.

[9] M. A. CREASY, "Limits for the ratio of means", *J.R Stat Soc*, B, Vol 16 (1954), pp. 186–194.

[10] R. A. FISHER, "The logic of inductive inference", *J R Stat Soc*, Vol 98 (1935), pp. 39–54.

[11] R A. FISHER, "Statistical methods and scientific induction", *J.R. Stat. Soc*, B, Vol. 17 (1955), 69–78.

[12] R A. FISHER, "On a test of significance in Pearson's Biometrika Table No 11", *J.R. Stat. Soc.*, B, Vol 18 (1956), pp 56–60.

[13] R. A. FISHER, *Statistical methods and scientific inference*. Edinburgh Oliver & Boyd, 1956.

[14] W. HOEFFDING, "The role of assumptions in statistical decisions", *Proc 3rd Berkeley Symp*, Vol. 1 (1955), pp. 105–114

[15] S. KARLIN AND H RUBIN, "Theory of decision procedures for distributions with monotone likelihood ratio", *Ann Math. Stat*, Vol. 27 (1956), pp 272–299

[16] O. KEMPTHORNE, "The randomization theory of experimental inference", *J. Am. Stat. Assn.* Vol. 50 (1955), pp. 946–967

[17] H. JEFFREYS, *The theory of probability*. Oxford, 2nd ed, 1946.

[18] E. L. LEHMANN AND H. SCHEFFE, "Completeness, similar regions and unbiased estimation, II, *Sankhya*, Vol. 15 (1955), pp 219–236.

[19] D. V. LINDLEY, "A statistical paradox", *Biometrika*, Vol. 44 (1957), pp. 187–192

[20] J. G. MAULDON, "Pivotal quantities for Wishart's and related distributions and a paradox in fiducial theory", *J.R Stat. Soc.*, B, Vol. 17 (1955), pp 79–85.

[21] J. NEYMAN, "Note on an article by Sir Ronald Fisher", *J.R Stat. Soc.*, B, Vol. 18 (1956), pp 288–294.

[22] J NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability", *Phil. Trans. Roy Soc*, A, Vol. 236 (1937), pp. 333–380.

[23] A. R. G. OWEN, "Ancillary statistics and fiducial distributions", *Sankhya*, Vol. 9, (1948), pp. 1–18.

[24] E. S. PEARSON, "Statistical concepts in their relation to reality", *J R Stat. Soc*, B, Vol. 17 (1955), pp. 204–207.

[25] J. PRZYBOROWSKI AND J. WILENSKI, "Homogeneity of results in testing samples from Poisson series", *Biometrika*, Vol. 31 (1939), pp. 313–323.

[26] C. Stein, "A property of some tests of composite hypotheses", *Ann. Math. Stat.*, Vol. 22 (1951), pp. 475–476.
[27] J. W. Tukey, "Fiducial inference", unpublished lectures.
[28] J. W. Tukey, "Some examples with fiducial relevance", *Ann. Math. Stat.*, Vol. 28 (1957), pp. 687–695.
[29] B. L. Welch, "On confidence limits and sufficiency with particular reference to parameters of location", *Ann. Math. Stat.*, Vol. 10 (1939), pp. 58–69.
[30] B. L. Welch, "Generalisation of Student's problem", *Biometrika*, Vol. 34 (1947), pp. 28–35.
[31] B. L. Welch, "Note on some criticisms made by Sir Ronald Fisher", *J.R. Stat. Soc.*, B, Vol. 18 (1956), pp. 297–302.

ASYMPTOTIC DISTRIBUTION OF STOCHASTIC APPROXIMATION PROCEDURES[1]

BY JEROME SACKS[2]

California Institute of Technology

1. Introduction. Beginning with the paper of Robbins and Monro [11] much work has been done in stochastic approximation. The Robbins-Monro procedure (see [11] or Section 3 below) for finding the root of a regression equation and the Kiefer-Wolfowitz procedure (see [9] or Section 4 below) for finding the maximum of a regression function have been the chief objects of investigation. The investigations that have been carried out on these procedures have been along two lines: the first being concerned with conditions under which the procedures (i.e., the sequence $\{X_n\}$ of approximating random variables) converge, in some sense, and the second being concerned with the speed of convergence and the asymptotic distribution of the procedures. For details concerning these investigations we refer the reader to the literature; some account of them may be found in Sections 3, 4, and 5 when they relate to the context. In particular the results relating to conditions for convergence are all subsumed in the work of Dvoretzky [7], Wolfowitz [12], and Block [1].

Chung [5] was the first to give any results about the asymptotic distribution of these procedures in his treatment of the Robbins-Monro procedure, and his methods (see the next paragraph) have been the basis for all work done heretofore in this direction. Hodges and Lehmann [8] improved some of Chung's results. Derman [6] used Chung's methods to obtain some results for the Kiefer-Wolfowitz procedure and Burkholder [4] extended Chung's methods to obtain further results on the asymptotic distribution of the Kiefer-Wolfowitz procedure.

Chung's method for obtaining his results on the asymptotic normality of the appropriately normalized sequence $\{X_n\}$ is to compute sufficiently fine estimates for the moments of $X_n - \theta$ (θ is the root of the regression equation) and then to apply the method of moments. As we noted above all previous work on the asymptotic distribution of the two procedures in question has been based on Chung's methods. The main feature of the present work is that we do away with the method of moments by, instead, utilizing a central limit theorem for dependent random variables and obtain more general and more complete results about the asymptotic normality of $\{X_n\}$ for both procedures by using a different method of proof—the method of proof we use may be seen by referring to that portion of the proof of Theorem 1 which lies between (3.8) and (3.9c). In addition, in Examples 1 and 2 in Section 4 we show that some of the results obtained here are best possible in a certain sense.

Received June 3, 1957; revised December 2, 1957

[1] This research was done in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Cornell University and was sponsored by the Office of Naval Research and the Air Force.

[2] Now at the Department of Mathematical Statistics, Columbia University

One of the complications arising from use of the method of moments is that the computations needed there are not feasible unless $\{a_n\}$ and $\{c_n\}$ (see (3.0) and (4.0)) are of the special type $a_n = an^{-t}$, $c_n = cn^{-r}$. While we take $a_n = an^{-1}$ in Sections 3, 4, and 5 (see Section 6 for further remarks on this choice of a_n) one of the reasons why we can obtain better results for the Kiefer-Wolfowitz procedure than heretofore obtained is that the method of proof we use permits a wider choice for $\{c_n\}$. Other desirable features of the method of proof presented here are that restrictions are needed only on the second moments of $Z(x)$ (the method of moments requires restrictions on all moments) and that the method can be used without difficulty on some multi-dimensional analogues (see Section 5) of the procedures.

In Section 3 we treat the Robbins-Monro procedure and in Section 4 we discuss the Kiefer-Wolfowitz procedure. Section 5 is devoted to some multi-dimensional analogues of the procedures. Section 6 discusses some further consequences and extensions of the results of earler sections. In Section 2 we collect some lemmas and computations which are used repeatedly in later sections.

The author would like to take this opportunity to acknowledge his debt to Professors J. Kiefer and J. Wolfowitz for their direction and assistance during the course of this research.

2. Preliminaries. In this section we will collect and prove several simple results which are used repeatedly in later sections. In addition, we will state and prove the central limit theorem which we use in succeeding sections. In what follows D_1, D_2, etc., will denote constants appropriately chosen to suit the context in which they appear.

Let $\{a_n\}$ be a sequence of positive real numbers such that

$$\sum a_n = \infty, \qquad \sum a_n^2 < \infty \tag{2.0}$$

Except for Lemma 1 it will always be assumed in this section that $a_n = an^{-1}$ for some $a > 0$. Let

$$\begin{aligned} \beta_{mn} &= \prod_{j=m+1}^{n} (1 - a_j) && 0 \leqq m < n \\ &= 1 && m = n \end{aligned} \tag{2.1}$$

It is well known that

$$(1 - \epsilon_m) \exp\left\{-\sum_{j=m+1}^{n} a_j\right\} \leqq \beta_{mn} \leqq (1 + \epsilon_m) \exp\left\{-\sum_{j=m+1}^{n} a_j\right\} \tag{2.2}$$

for all $n \geqq m$, where $\epsilon_m \to 0$ as $m \to \infty$. In particular, if, for $a > 0$, $a_n = an^{-1}$ we have

$$(1 - \epsilon_m')m^a n^{-a} \leqq \beta_{mn} \leqq (1 + \epsilon_m')m^a n^{-a} \tag{2.3}$$

where $\epsilon_m' \to 0$ as $m \to \infty$.

LEMMA 1. *Let $\{W_m\}$ be a sequence of real numbers converging to W where W may*

be taken to be ∞. *Then, for any positive integer* m_0,

$$\lim_{n\to\infty} \sum_{m=m_0}^{n} a_m \beta_{mn} W_m = W$$

PROOF. For any fixed m it follows from (2.2) that $\lim_{n\to\infty} \beta_{mn} = 0$. Since $a_m\beta_{mn} = \beta_{mn} - \beta_{m-1,n}$ we have, for any fixed m_1,

$$\lim_{n\to\infty} \sum_{m=m_1}^{n} a_m \beta_{mn} = \lim_{n\to\infty} (1 - \beta_{m_1-1,n}) = 1.$$

The conclusion of Lemma 1 now follows quite easily.

Let $\{c_m\}$ be a sequence of positive real numbers. For each n let

$$h_n = (\textstyle\sum_{m=1}^{n} a^2 c_m^{-2} m^{-2} \beta_{mn}^2)^{-1/2}.$$

LEMMA 2. *Let* $a > 1/2$ *and suppose that* $c_m \leqq c < \infty$ *for all* m. *Then, if* m_0 *is some fixed positive integer*

$$\lim_{n\to\infty} h_n \beta_{m_0 n} = 0.$$

PROOF. If $m_0 > a - 1$ let $m_1 = m_0$; otherwise, let m_1 be the smallest integer greater than $a - 1$. To prove the lemma it is obviously sufficient to prove that $h_n\beta_{m_1 n} \to 0$ as $n \to \infty$. Using (2.2) we obtain

$$\begin{aligned} h_n^2 \beta_{m_1 n}^2 &\leqq D_1 h_n^2 n^{-2a} \leqq D_2 n^{-2a} \left(\sum_{m=m_1}^{n} a m^{-1} \beta_{mn} c_m^{-2} m^{a-1} n^{-a} \right)^{-1} \\ &= D_2 \left(\sum_{m=m_1}^{n} a m^{-1} \beta_{mn} c_m^{-2} m^{a-1} n^{a} \right)^{-1} \end{aligned} \tag{2.4}$$

If $a > 1/2$ then $n^a m^{a-1} c_m^{-2} > c^{-2} m^{2a-1}$ which goes to ∞ as $m \to \infty$ and hence, by Lemma 1, the last term in (2.4) goes to 0 as $n \to \infty$.

LEMMA 3. *Let* $a > 1/2$ *and suppose that* $c_m \leqq c < \infty$ *for all* m. *Let* $\{W_m\}$ *be a sequence of real numbers converging to* W *where* W *may be* ∞. *Then, if* m_0 *is a fixed positive integer*

$$\lim_{n\to\infty} h_n^2 \sum_{m=m_0}^{n} a^2 c_m^{-2} m^{-2} \beta_{mn}^2 W_m = W.$$

PROOF. The proof is easily accomplished upon noting that, for any fixed m_1

$$\lim_{n\to\infty} h_n^2 \sum_{m=m_1}^{n} a^2 m^{-2} c_m^{-2} \beta_{mn}^2 = 1.$$

LEMMA 4. *Let* $\{d_m\}$ *be a sequence of positive numbers such that*

$$d_m d_{m+1}^{-1} = 1 + \epsilon_m m^{-1} \qquad \textit{where } \epsilon_m \to 0 \textit{ as } m \to \infty. \tag{2.5}$$

Let $q > -1$. *Then, for any positive integer* m_0,

$$\sum_{m=m_0}^{n} d_m m^q \sim (1 + q)^{-1} d_n n^{q+1}$$

Examples of sequences satisfying (2.5) are easy to obtain. For example, $d_m = (\log m)^u$ satisfies (2.5) for all real u. Note also that if $\{d_m\}$ satisfies (2.5) then $\{d_m^\rho\}$ also satisfies (2.5) for any real number ρ.

PROOF. Since $\sum_{m=1}^n m^q \sim (1+q)^{-1}n^{q+1}$ what we have to show will be accomplished if we show that

$$\frac{\sum_{m=1}^n d_m m^q}{d_n \sum_{m=1}^n m^q} - 1 = \frac{\sum_{m=1}^{n-1} (d_m - d_{m+1}) \sum_{j=1}^m j^q + d_n \sum_{j=1}^n j^q}{d_n \sum_{m=1}^n m^q} - 1$$

$$= \frac{\sum_{m=1}^{n-1} d_m d_n^{-1}(d_m d_{m+1}^{-1} - 1) \sum_{j=1}^m j^q}{\sum_1^n m^q} = A_n \quad \text{(say)}$$

goes to 0 as $n \to \infty$. Using (2.5) we see that for n sufficiently large

$$d_n = d_1 \prod_{j=1}^{n-1} d_{j+1} d_j^{-1} \geqq D_3 e^{\sum_{j=1}^{n-1} \epsilon_j j^{-1}} \geqq D_4 e^{-\epsilon \sum_{j=m_1}^{n-1} j^{-1}} \geqq D_5 m_1^\epsilon n^{-\epsilon}$$

where m_1 is chosen so that $|\epsilon_j| \leqq \epsilon < q + 1$ for all $j \geqq m_1$. Thus $d_n n^{q+1} \to \infty$ as $n \to \infty$ and, therefore, in order to show that $A_n \to 0$, we can start the outer sum in the numerator of A_n at $m = m_1$.

By use of (2.5) we have, for $n > m \geqq m_1$,

$$d_m d_n^{-1} = \prod_{j=m}^{n-1} d_j d_{j+1}^{-1} \leqq \prod_{j=m}^{n-1} (1 + \epsilon j^{-1}) \leqq D_6 n^\epsilon m^{-\epsilon}$$

and

$$|d_m d_{m+1}^{-1} - 1| \leqq \epsilon m^{-1}$$

That A_n must go to 0 now follows because for all n

$$\frac{\sum_{m=m_1}^n D_6 n^\epsilon m^{-\epsilon} \epsilon m^{-1} \sum_{j=1}^m j^q}{\sum_{m=1}^n m_q} \leqq \epsilon D_7 n^{\epsilon - q - 1} \sum_{m_1}^n m^{-\epsilon-1} m^{q+1} \leqq \epsilon D_8$$

and because ϵ is arbitrary.

LEMMA 5. *Let $c_m = d_m m^{-r}$ where $r \geqq 0$ and where $\{d_m\}$ satisfies (2.5). Let a be a real number greater than 1/2. Then, for any positive integer m_0, and any positive number ρ,*

$$\sum_{m=m_0}^n a^2 c_m^{-\rho} m^{-2} \beta_{mn}^2 \sim a^2(2a + r\rho - 1)^{-1} c_n^{-\rho} n^{-1} \tag{2.6}$$

as $n \to \infty$. In particular, if $m_0 = 1$ and $\rho = 2$ (2.6) becomes

$$h_n^2 \sim a^{-2}(2a + 2r - 1) n c_n^2 \tag{2.7}$$

PROOF. Let $\epsilon > 0$ and let m_1 be large enough so that in (2.3) $\epsilon_m' < \epsilon$ for $m > m_1$. Then, using (2.3) and Lemma 4—the conditions of Lemma 4 are satisfied if one takes into account the conditions stated here and the remarks following the

statement of Lemma 4—we obtain

$$(2.8)\quad \begin{aligned} a^2 \sum_{m=m_0}^{n} c_m^{-\rho} m^{-2} \beta_{mn}^2 &\leqq (1+\epsilon)a^2 \sum_{m_1}^{n} c_m^{-\rho} m^{2a-2} n^{-2a} + O(n^{-2a}) \\ &= (1+\epsilon)a^2 \sum_{m_1}^{n} d_m^{-\rho} m^{2a+\rho r-2} n^{-2a} + O(n^{-2a}) \\ &\leqq (1+\alpha_n)a^2(2a+\rho r-1)^{-1} d_n^{-\rho} n^{2a+\rho r-1} n^{-2a} + O(n^{-2a}) \\ &= (1+\alpha_n)a^2(2a+\rho r-1)^{-1} c_n^{-\rho} n^{-1} + O(n^{-2a}) \end{aligned}$$

where $\alpha_n \to \epsilon$ as $n \to \infty$. Similar calculation produces

$$a^2 \sum_{m_0}^{n} c_m^{-\rho} m^{-2} \beta_{mn}^2 \geqq (1+\alpha_n)a^2(2a+\rho r-1)^{-1} c_n^{-\rho} n^{-1}$$

Since $n^{1-2a} c_n^{\rho} \to 0$ as $n \to \infty$ and since ϵ is arbitrary we have achieved the desired result.

We shall now state and prove a central limit theorem which we use later in an essential way. The multi-dimensional version we give (see Lemma 6) is a direct generalization of the one-dimensional result which may be found in Loeve [10], p. 377 C. The proof we give is likewise a direct generalization of the proof given in [10]. In Sections 3 and 4 it will suffice to consider only the one-dimensional case; we make use of the result for higher dimensions in Section 5.

With all vectors considered as elements of q-dimensional Euclidean space we adopt the following notation. If x, y are vectors $[x, y]$ will denote their inner product. The norm of a vector x we denote by $|x|$ and, of course, is equal to $[x, x]^{1/2}$. If B is a $q \times q$ matrix we define in the usual way,

$$\| B \| = \sup_{|x|=1} [Bx, Bx]^{1/2}$$

The obvious facts that $|Bx| \leqq \|B\| \, |x|$ and that $\|B_1 B_2\| \leqq \|B_1\| \, \|B_2\|$ will be useful below. I will denote the identity $q \times q$ matrix. B' and x' will denote the transposes of the matrix B and vector x respectively. Unless otherwise indicated a vector is to be considered a column vector.

Let $\{U_{nk};\, 1 \leqq k \leqq n,\, n \geqq 1\}$ be a family of vector random variables, the distribution of U_{nk} being denoted by F_{nk}. Let $V_{nk} = (U_{n1}, \cdots, U_{n,k-1})$ and suppose that $E(U_{nk} \mid V_{nk}) = 0$ with probability one. Denote the covariance matrix of U_{nk} by s_{nk} i.e., $s_{nk} = E(U_{nk} U'_{nk})$. Let $r_{nk} = E(U_{nk} U'_{nk} \mid V_{nk})$. Let $U_n = \sum U_{nk}$, $s_n = \sum s_{nk}$, and $r_n = \sum r_{nk}$ where all three summations are over $1 \leqq k \leqq n$. For $\epsilon > 0$ define $\phi_{nk}^{\epsilon} = 1$ if $|U_{nk}| > \epsilon$, $\phi_{nk}^{\epsilon} = 0$ otherwise.

LEMMA 6. *If*

$$(2.9)\quad \lim_{n\to\infty} \sum_{k=1}^{n} E(\| r_{nk} - s_{nk} \|) = 0$$

$$(2.10)\quad \sup_{n} \sum_{k=1}^{n} E(| U_{nk} |^2) < \infty,$$

and, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{k=1}^{n} E(| U_{nk} |^2 \phi_{nk}^{\epsilon}) = 0, \tag{2.11}$$

and if $s_n \to s$, i.e., $\|s_n - s\| \to 0$, *then* U_n *is asymptotically normal with mean* 0 *and covariance matrix* s.

PROOF. Let F and G be q-dimensional distribution functions with characteristic functions f and g and finite covariance matrices C and D respectively, and let $H = F - G$. Let θ_1, θ_2 denote quantities whose absolute value is less than 1. Let $\Delta = \{x | |x| \leqq \epsilon\}$ and let Δ' be the complement of Δ. Then, for fixed t and $\epsilon < 1/|t|$,

$$\begin{aligned} | f(t) - g(t) | &\leqq \left| \int [t, x] \, dH(x) \right| + \frac{1}{2} \left| \int_{\Delta} [t, x]^2 \, dH(x) \right| \\ &\quad + \left| \int_{\Delta} \theta_1 [t, x]^3 \, dH(x) \right| + \left| \int_{\Delta'} \theta_2 [t, x]^2 \, dH(x) \right| \\ &\leqq \left| \int [t, x] \, dH(x) \right| + \left| \int [t, x]^2 \, dH(x) \right| \\ &\quad + \epsilon | t | \int [t, x]^2 \, d(F + G) + 3 \int_{\Delta'} [t, x]^2 \, d(F + G) \\ &\leqq \left| \int [t, x] \, dH(x) \right| + | t |^2 \| C - D \| \\ &\quad + \epsilon | t |^3 \int | x |^2 \, d(F + G) + 3 | t |^2 \int_{\Delta'} | x |^2 \, d(F + G) \end{aligned} \tag{2.12}$$

Let G_{nk} denote the normal distribution with mean 0 and covariance matrix s_{nk}. Let $\{Y_{nk} ; 1 \leqq k \leqq n, n \geqq 1\}$ be a family of independent random variables with the distribution of Y_{nk} being G_{nk}. In addition, take $\{Y_{nk}\}$ to be independent of $\{U_{nk}\}$. It is easy to see that $Y_n = Y_{n1} + \cdots + Y_{nn}$ is asymptotically normal with mean 0 and covariance matrix s.

Let f_{nk}, f_n, g_{nk}, and g_n denote the characteristic functions of U_{nk}, U_n, Y_{nk}, and Y_n respectively. Let

$$f_{nk}^*(t) = E(e^{i[t, U_{nk}]} \mid V_{nk}).$$

To prove the lemma it is clearly sufficient to prove that, for each fixed t,

$$\lim_{n\to\infty} |f_n(t) - g_n(t)| = 0.$$

Let $W_{nk} = U_{n1} + \cdots + U_{n,k-1} + Y_{n,k+1} + \cdots + Y_{nn}$ for $1 < k < n$, $W_{n1} = Y_{n2} + \cdots + Y_{nn}$, $W_{nn} = U_{n1} + \cdots + U_{n,n-1}$. Then

$$\begin{aligned} | f_n(t) - g_n(t) | &= | E(e^{i[t, U_n]} - e^{i[t, Y_n]}) | \\ &= \left| E \sum_{k=1}^{n} (e^{i[t, U_{nk}]} - e^{i[t, Y_{nk}]}) e^{i[t, W_{nk}]} \right| \end{aligned} \tag{2.13}$$

$$\leqq \sum_{k=1}^{n} E\,|\,f_{nk}^{*}(t) - g_{nk}(t)\,|$$

From (2.12), (2.13), and the fact that $E(U_{nk} \mid V_{nk}) = 0$ we obtain

$$\begin{aligned} |\,f_n(t) - g_n(t)\,| \leqq |\,t\,|^2 \sum_{k=1}^{n} E\,\|\,r_{nk} - s_{nk}\,\| + 2\epsilon\,|\,t\,|^3 \sum_{k=1}^{n} E\,|\,U_{nk}\,|^2 \\ + 3\,|\,t\,|^2 \sum_{k=1}^{n} E(\phi_{nk}^{\epsilon}\,|\,U_{nk}\,|^2) + 3\,|\,t\,|^2 \sum_{k=1}^{n} \int_{\Delta'} |\,x\,|^2\, dG_{nk} \end{aligned} \tag{2.14}$$

As $n \to \infty$ the first and third terms on the right-hand side of (2.14) go to 0 because of (2.9) and (2.11), the second term is $O(\epsilon)$ because of (2.10), and the last term goes to 0 because G_{nk} is normal with covariance matrix s_{nk} and $\|s_{nk}\|$ goes to 0 as $n \to \infty$ uniformly in $k \leqq n$. Since ϵ is arbitrary this finishes the proof of the lemma.

3. The Robbins-Monro Procedure. Let M be a fixed function such that the equation $M(x) = \alpha$ has a unique solution $x = \theta$. For each x let $Y(x)$ be a random variable with $EY(x) = M(x)$. The Robbins-Monro procedure for "finding" θ is defined as follows. Let $\{a_n, n > 0\}$ be a sequence of positive numbers such that

$$\sum a_n = \infty, \qquad \sum a_n^2 < \infty. \tag{3.0}$$

Let X_1 be some fixed number (X_1 may be taken to be an arbitrary random variable for what follows since, if $EX_1^2 < \infty$, the same proofs will hold, while, if $EX_1^2 = \infty$, the results are obtained by truncating X_1 and using the results for the case $EX_1^2 < \infty$) and define $\{X_n, n > 1\}$ by the recursion

$$X_{n+1} = X_n - a_n(Y(X_n) - \alpha) \tag{3.1}$$

where $Y(X_n)$ is a random variable whose conditional distribution given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Y(x_n)$. Letting $Z(x) = Y(x) - M(x)$ (3.1) becomes

$$X_{n+1} = X_n - a_n[M(X_n) - \alpha + Z(X_n)], \tag{3.2}$$

$EZ(x) = 0$ for all x, and the conditional distribution of $Z(X_n)$ given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Z(x_n)$. We note for future use that, as a consequence of this,

$$E(Z(X_n) \mid Z(X_1), \cdots, Z(X_{n-1})) = 0 \tag{3.3}$$

with probability one.

For Theorem 1 we make the following assumptions about $M(x)$ and $Z(x)$. The connection between these assumptions and those made by previous authors is pointed out below.

Assumption (A1). M is a Borel-measurable function; $M(\theta) = \alpha$ and

$$(x - \theta)(M(x) - \alpha) > 0$$

for all $x \neq \theta$.

ASSUMPTION (A2). For some positive constants K and K_1, and for all x

$$K \mid x - \theta \mid \leqq |M(x) - \alpha| \leqq K_1 \mid x - \theta \mid$$

ASSUMPTION (A3). For all x

$$M(x) = \alpha + \alpha_1(x - \theta) + \delta(x, \theta)$$

where $\delta(x, \theta) = o(|x - \theta|)$ as $x - \theta \to 0$ and where $\alpha_1 > 0$.

ASSUMPTION (A4).

$$\text{(a)} \sup_x EZ^2(x) < \infty; \qquad \text{(b)} \lim_{x \to \theta} EZ^2(x) = \sigma^2$$

ASSUMPTION (A5).

$$\lim_{R\to\infty} \lim_{\epsilon\to 0^+} \sup_{|x-\theta|<\epsilon} \int_{\{|Z(x)|>R\}} Z^2(x)\, dP = 0$$

When $X_n \to \theta$ with probability one (for example, under (A1), (A2′), and (a) of (A4) as shown by Blum [2], and when (a) of (A4) holds, (A5) implies

$$\lim_{R\to\infty} \sup_k \int_{\{|Z(X_k)|>R\}} Z^2(X_k)\, dP = 0 \tag{3.4}$$

which is actually what is used in the proof below. The reason we state (A5) in the way we have is that it appears as a more natural condition than (3.4). Simple conditions which imply (A5) are given by

$$\{Z(x)\} \text{ are identically distributed} \tag{3.5}$$

or

$$\sup_{|x-\theta|<\epsilon} E\mid Z(x) \mid^{2+v} < \infty \tag{3.6}$$

for some $\epsilon > 0$ and some $v > 0$.

Assumption (A2) can be weakened to

ASSUMPTION (A2′). For all x and some positive constant K_1

$$|M(x) - \alpha| \leqq K_1 \mid x - \theta|$$

and, for every t_1, t_2 such that $0 < t_1 < t_2 < \infty$,

$$\inf_{t_1 \leqq |x-\theta| \leqq t_2} \mid M(x) - \alpha \mid > 0.$$

In Theorem 1′ we obtain the result of Theorem 1 with (A2′) replacing (A2)—the truncation device used in the proof there is due to Hodges and Lehmann [8].

Under (A1), (A2), (A3), the assumption that $EZ^2(x) = \sigma^2$ for all x, and the assumption that (3.6) hold for $\epsilon = \infty$ and all v Chung obtained the result of Theorem 1 (this is what is referred to in [5] as the "second case"). Hodges and Lehmann proved the result of Theorem 1 under (A1), (A2′), (A3), (A4), and the assumption that (3.6) hold for some $\epsilon > 0$ and all v. Thus Theorem 1′ includes these earlier results by virtue of the greater generality of (A5) over related conditions made by previous authors.

As before D_1, D_2, etc., will denote positive constants appropriately chosen for the context in which they appear.

THEOREM 1. *Suppose that Assumptions* (A1) *through* (A5) *are satisfied. Let* $a_n = An^{-1}$ *for* $n > 0$ *where* A *is such that* $2KA > 1$. *Then* $n^{1/2}(X_n - \theta)$ *is asymptotically normally distributed with mean* 0 *and variance* $A^2\sigma^2(2A\alpha_1 - 1)^{-1}$.

PROOF. There is no loss in assuming that $\alpha = \theta = 0$. Abbreviating $\delta(X_n, 0)$, $M(X_n)$ and $Z(X_n)$ by δ_n, M_n, and Z_n respectively, and using (A3) we rewrite (3.2) and obtain

$$X_{n+1} = (1 - A\alpha_1 n^{-1})X_n - An^{-1}\delta_n - An^{-1}Z_n \tag{3.7}$$

Let $a = A\alpha_1$ and let β_{mn} be as in (2.1) with the a_j in (2.1) replaced by aj^{-1}. Iteration of (3.7) then yields

$$X_{n+1} = \beta_{on}X_1 - A\sum_{m=1}^{n} m^{-1}\beta_{mn}\delta_m - A\sum_{m=1}^{n} m^{-1}\beta_{mn}Z_m \tag{3.8}$$

Let $h_n = \left(\sum_{m=1}^{n} a^2m^{-2}\beta_{mn}^2\right)^{-1/2}$. Then, by Lemma 5,

$$h_n \sim (2a - 1)^{1/2}a^{-1}n^{1/2}$$

Hence, proving that $n^{1/2}X_n$ is asymptotically normal with mean 0 and variance $A^2\sigma^2(2a - 1)^{-1}$ is equivalent to proving

(3.9) h_nX_n is asymptotically normal with mean 0 and variance $A^2\sigma^2a^{-2}$.

Using (3.8) it is clear that we can show (3.9) by proving

$$h_n\beta_{on} \to 0 \tag{3.9a}$$

$$h_n\sum_{m=1}^{n} am^{-1}\beta_{mn}\delta_m \to 0 \text{ in probability} \tag{3.9b}$$

(3.9c) $h_n\sum_{m=1}^{n} am^{-1}\beta_{mn}Z_m$ is asymptotically normal with mean 0 and variance σ^2.

(3.9a) follows immediately from Lemma 2 with $c_m = 1$ for all m. To prove (3.9c) we will invoke Lemma 6 with $q = 1$ and $U_{nk} = h_nak^{-1}\beta_{kn}Z_k$. To see that we can do so observe first that by (3.3)

$$E(U_{nk} \mid U_{n1}, \cdots, U_{n,k-1}) = E(U_{nk} \mid Z_1, \cdots, Z_{k-1}) = 0.$$

Let $\phi_{nk} = 1$ if $|U_{nk}| \geqq \epsilon$ and $\phi_{nk} = 0$ otherwise, and observe that in order to verify (2.11) we have to check that $\sum_{k=1}^{n} E(\phi_{nk}U_{nk}^2) \to 0$ or, what is the same,

$$h_n^2\sum_{k=1}^{n} a^2k^{-2}\beta_{kn}^2E(\phi_{nk}Z_k^2) \to 0 \tag{3.10}$$

Noticing, by Lemma 5 and (2.3) that $\phi_{nk} = 1$ implies, for some $\epsilon' > 0$, that $|Z_k| \geqq \epsilon'n^{a-1/2}k^{1-a} \geqq \epsilon'k^{1/2}$, we apply (3.4) which is obtained from (A5) and obtain

$$\lim_{k\to\infty} E(\phi_k'Z_k^2) = 0 \tag{3.11}$$

where $\phi_k' = 1$ if $|Z_k| \geqq \epsilon' k^{1/2}$ and $\phi_k' = 0$ otherwise. Since $\phi_k' \geqq \phi_{nk}$, applying Lemma 3 with $c_m = 1$ for all m and using (3.11) shows that (3.10) is valid. Verifying (2.9) is equivalent to showing

$$\lim_{n\to\infty} h_n^2 \sum_{k=1}^{n} a^2 k^{-2} \beta_{kn}^2 E \mid E'[Z^2(X_k)] - EZ^2(X_k) \mid = 0 \tag{3.12}$$

where E' denotes conditional expected value with the conditioning being by V_{nk}. Use again of Lemma 3 shows that it is sufficient to prove

$$\lim_{k\to\infty} E \mid E'[Z^2(X_k)] - EZ^2(X_k) \mid = 0 \tag{3.13}$$

But (3.13) follows easily by observing that the expression between the absolute value signs is uniformly bounded ((a) of (A4)) so that Lebesgue's theorem is applicable, and by observing that (b) of (A4) together with the convergence of X_k to θ w.p.1 imply

$$\lim_{k\to\infty} E'[Z^2(X_k)] = \lim_{k\to\infty} EZ^2(X_k) = \sigma^2. \tag{3.14}$$

(3.14) and Lemma 3 also serve to show that (2.10) is satisfied with

$$\lim_{n\to\infty} s_n = \sigma^2 \tag{3.15}$$

This completes the verification that Lemma 6 is applicable and therefore establishes (3.9c).

To prove (3.9b) we require the estimate that $EX_n^2 = O(n^{-1})$. This estimate is obtained by Chung [5] but we obtain it here for completeness. The methods are essentially the same.

Squaring both sides of (3.2), taking expected values, and using (A4) we get

$$EX_{n+1}^2 = E(X_n - An^{-1}M_n)^2 + O(n^{-2}) \tag{3.16}$$

Then, by (A1) and (A2), for ϵ sufficiently small so that $2KA - \epsilon > 1$, and for n sufficiently large, say $n > N_1$,

$$\begin{aligned} EX_{n+1}^2 &\leqq (1 - 2KAn^{-1} + A^2K_1^2n^{-2})EX_n^2 + O(n^{-2}) \\ &\leqq (1 - (2KA - \epsilon)n^{-1})EX_n^2 + D_1n^{-2} \end{aligned} \tag{3.17}$$

Let $p = 2KA - \epsilon$ and let β_{mn}' be defined by (2.1) with $a_j = pj^{-1}$. Choose N_1 large enough so that $p < N_1$ (this is to guarantee that $\beta_{mn}' > 0$ for $m \geqq N_1$ so that (3.18) can hold). Iteration of (3.17) yields

$$\begin{aligned} EX_{n+1}^2 &\leqq D_1 \sum_{m=N_1+1}^{n} m^{-2}\beta_{mn}' + \beta_{N_1 n}' EX_{N_1+1}^2 \leqq D_2 n^{-1} + D_3 n^{-p} \\ &= O(n^{-1}) \end{aligned} \tag{3.18}$$

which is the estimate we require.

Let $t > 0$. Since $\delta(x) = o(|x|)$, for $t > 0$ we can find $\epsilon > 0$ with the property

that

(3.19) $$|\delta(x)| \leqq t^2|x| \qquad \text{for } |x| \leqq \epsilon.$$

As was pointed out above $X_n \to 0$ w.p 1; hence, we can choose N_2 so that

(3.20) $$P\{|X_j| \leqq \epsilon, j \geqq N_2\} > 1 - t.$$

Let N_3 be larger than N_1 and N_2 and such that $a < N_3 + 1$. Then, denoting $h_n\sum_{m=N_3}^{n} am^{-1}\beta_{mn}\delta_m$ by V_n and $h_n \sum_{m=N_3}^{n} am^{-1}\beta_{mn} |X_m|$ by V_n^*, and using (3.20), (3.19), a Chebyshev-type inequality, (3.18), and Lyapounov's inequality, and (2.3) we have for $n > N_3$,

(3.21) $$\begin{aligned} P\{|V_n| > t\} &\leqq t + P\{|V_n| > t, |X_j| \leqq \epsilon, j \geqq N_3\} \\ &\leqq t + P\{t^2V_n^* > t\} \leqq t + tEV_n^* \\ &\leqq t + D_4th_n \sum_{N_3}^{n} m^{-1}\beta_{mn} m^{-1/2} \leqq D_6 t. \end{aligned}$$

(3.21) together with the fact that $h_n\beta_{mn} \to 0$ for any fixed m (Lemma 2) establishes (3.9b) and finishes the proof of the theorem.

THEOREM 1'. *Suppose that Assumptions* (A1), (A2'), (A3), (A4), *and* (A5) *are satisfied. Let $a_n = An^{-1}$ where A is such that $A\alpha_1 > 1/2$. Then $n^{1/2}(X_n - \theta)$ is asymptotically normally distributed with mean* 0 *and variance $A^2\sigma^2(2A\alpha_1 - 1)^{-1}$.*

PROOF. We assume with no loss of generality that $\alpha = \theta = 0$. Let $t > 0$ be such that $A(\alpha_1 - t) > 1/2$. Let $K = \alpha_1 - t$. Then we can find an $\epsilon > 0$ such that for $|x| \leqq \epsilon$

(3.22) $$K|x| \leqq |M(x)| \leqq K_1|x|.$$

Define $M'(x) = M(x)$ if $|x| \leqq \epsilon$, $M'(x) = Kx$ if $|x| > \epsilon$.

Since under (A1), (A2'), and (A4), $X_n \to 0$ w.p.1 we can find N so that for $u > 0$,

(3.23) $$P\{|X_j| \leqq \epsilon, j \geqq N\} > 1 - u.$$

Let $X_1' = X_{N+1}$ and define $\{X_n', n \geqq 1\}$ by the recursion

(3.24) $$X_{n+1}' = X_n' - a_{n+N}M'(X_n') - a_{n+N}Z(X_n')$$

It is clear that the assumptions of Theorem 1' together with (3.22) show that Theorem 1 is applicable to X_n', M', $\{a_{n+N}\}$. Hence, for all y,

(3.25) $$\lim_{n\to\infty} P\{(N + n)^{1/2}X_{n+1}' < y\} = F(y)$$

where F is the normal distribution function with mean 0 and variance $A^2\sigma^2(2A\alpha_1 - 1)^{-1}$. Using (3.23) and (3.25) we obtain

$$\begin{aligned} \overline{\lim_{n\to\infty}} P\{n^{1/2}X_n < y\} &= \overline{\lim_{n\to\infty}} P\{(n + N)^{1/2}X_{n+N} < y\} \\ &\leqq \lim_{n\to\infty} P\{(n + N)^{1/2}(X_{n+N} - X_n') \end{aligned}$$

$$+ (n + N)^{1/2}X_n' < y; |X_j| \leqq \epsilon, j \geqq N\} + u \tag{3.26}$$
$$= \lim_{n\to\infty} P\{(n + N)^{1/2}X_n' < y\} + u$$
$$= F(y) + u$$

Similarly, we obtain

$$\underline{\lim}_{n\to\infty} P\{n^{1/2}X_n < y\} \geqq F(y) - u \tag{3.27}$$

Since u and y are arbitrary putting (3.26) and (3.27) together finishes the proof of the theorem.

4. The Kiefer-Wolfowitz Procedure. Let M be a fixed function with a unique maximum at $x = \theta$ (by making the obvious alterations in what follows we can replace "maximum" by "minimum"). For each x let $Y(x)$ be a random variable with $EY(x) = M(x)$. The Kiefer-Wolfowitz procedure for locating the maximum is defined as follows. Let $\{a_n\}$, $\{c_n\}$ be two sequences of positive numbers such that

$$\sum a_n = \infty, \qquad c_n \to 0, \qquad \sum a_n^2 c_n^{-2} < \infty \tag{4.0}$$

Let X_1 be a fixed number (by the same reasoning as in Sections 3 X_1 can be taken to be an arbitrary random variable for what follows) and define $\{X_n, n \geqq 2\}$ by the recursion

$$X_{n+1} = X_n - a_n c_n^{-1}[Y(X_n - c_n) - Y(X_n + c_n)] \tag{4.1}$$

where $Y(X_n \pm c_n)$ is a random variable whose conditional distribution given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Y(x_n \pm c_n)$. It is usually assumed that $Y(X_n - c_n)$ and $Y(X_n + c_n)$ are conditionally independent i.e., for all Borel sets A and B $P\{Y(X_n + c_n) \varepsilon A, Y(X_n - c_n) \varepsilon B \mid X_n\} = P\{Y(X_n + c_n) \varepsilon A \mid X_n\} P\{Y(X_n - c_n) \varepsilon B \mid X_n\}$. Though this is commonly the case in practice we do not make this assumption since it is unnecessary to do so. Whatever assumptions we do need to make about the joint distribution of $Y(X_n - c_n)$ and $Y(X_n + c_n)$ are contained in (B5). Letting $Z(x) = Y(x) - M(x)$ and writing M_n for $M(X_n - c_n) - M(X_n + c_n)$ and Z_n for $Z(X_n - c_n) - Z(X_n + c_n)$, (4.1) becomes

$$X_{n+1} = X_n - a_n c_n^{-1}(M_n + Z_n), \tag{4.2}$$

$EZ(x) = 0$ for all x, and the conditional distribution of Z_n given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Z(x_n - c_n) - Z(x_n + c_n)$. We note that, as a consequence of this,

$$E(Z_n \mid Z_1, \cdots, Z_{n-1}) = E(Z_n \mid X_1, \cdots, X_n) = 0 \tag{4.3}$$

with probability one.

We now make the assumptions we require for Theorem 2. Other assumptions relevant to later theorems are listed further on. The connection between these

assumptions and those made by previous authors is pointed out below. As before D_1, D_2, etc. will denote appropriately chosen positive constants.

ASSUMPTION (B1). $M(x)$ is a Borel-measurable function, has a unique maximum at $x = \theta$, and, for $0 < t_0 < t_1 < t_2 < \infty$,

$$\inf_{\substack{t_1 \leqq |x-\theta| \leqq t_2 \\ 0<\epsilon\leqq t_0}} \frac{(x-\theta)(M(x-\epsilon) - M(x+\epsilon))}{\epsilon} > 0 \tag{4.4}$$

In addition, for all x and suitable D_1 and D_2,

$$|M(x+1) - M(x)| < D_1 + D_2|x| \tag{4.5}$$

ASSUMPTION (B2). For all x

$$M(x) = \alpha_0 - \alpha(x-\theta)^2 + \delta(x, \theta)$$

where α_0 is some real number, $\alpha > 0$, and $\delta(x, \theta) = o(|x - \theta|^2)$ as $x - \theta \to 0$

ASSUMPTION (B3). For some $c_0 > 0$ there exist positive constants K_1 and K_2 such that, for all x and all c for which $0 < c \leqq c_0$,

$$K_1(x-\theta)^2 \leqq (x-\theta)[M(x-c) - M(x+c)]c^{-1} \leqq K_2(x-\theta)^2$$

ASSUMPTION (B4). For every $\epsilon > 0$ there exists $c_\epsilon > 0$ such that, for all c satisfying $0 < c \leqq c_\epsilon$ and all x satisfying $|x - \theta| < c$,

$$|\delta(x-c, \theta) - \delta(x+c, \theta)|c^{-1} \leqq \epsilon|x-\theta|$$

ASSUMPTION (B5).

$$\sup_x EZ^2(x) = s < \infty \tag{4.6}$$

$$\lim_{\substack{x\to\theta \\ a\to 0}} E[Z(x-a) - Z(x+a)]^2 = \sigma^2. \tag{4.7}$$

In case $Z(X_m - c_m)$ and $Z(X_m + c_m)$ are uncorrelated we can replace (4.7) by

$$\lim_{x\to\theta} EZ^2(x) = \sigma^2/2 \tag{4.8}$$

ASSUMPTION (B6).

$$\lim_{R\to\infty} \lim_{\epsilon\to 0^+} \sup_{|x|<\epsilon} \int_{\{|Z(x)|>R\}} Z^2(x)\, dP = 0$$

When $X_n \to \theta$ w.p.1 (for example, under (B1) and (4.6) as shown by Burkholder [4] and Dvoretzky [7]—Blum [2] proved convergence w.p.1 earlier but under stronger restrictions) (B6) implies

$$\lim_{R\to\infty} \sup_k \int_{\{|Z_k|>R\}} Z_k^2\, dP = 0 \tag{4.9}$$

The remarks made about (3.4) and (A5) pertain here to (4.9) and (B6) and, as with (A5), (B6) is satisfied if either (3.5) or (3.6) is fulfilled.

In Theorem 2′ we obtain the same result as in Theorem 2 with (B3) replaced by the weaker restriction

ASSUMPTION (B3′). For some $c_0 > 0$ there exist positive constants K_1 and K_2 such that, for all x in some neighborhood of θ and all c for which $0 < c \leqq c_0$,

$$K_1(x - \theta)^2 \leqq (x - \theta)[M(x - c) - M(x + c)]c^{-1} \leqq K_2(x - \theta)^2$$

(B3) ((B3′)) is used only for Theorem 2 (2′); it is replaced by a different condition for later theorems. (B4) which is also used only for Theorems 2 and 2′ is fulfilled whenever M satisfies (B2), (B3′), and has a continuous second derivative in some neighborhood of θ with $M''(\theta) = -2\alpha$ (i.e., $\delta''(\theta) = 0$). When (B2), (B3′), and (B4) hold simultaneously it is redundant to require the lower inequality in (B3′).

It is easy to see that (B3) (also (B3′)) implies that M is symmetric in some neighborhood of θ; in fact, $M(\theta - c) = M(\theta + c)$ for all $c < c_0$. If (B3) is satisfied and the interval of symmetry is known, i.e., c_0 is known, Burkholder was able to show that modifying the Kiefer-Wolfowitz procedure by taking $c_n = c_0$ for all n will yield, under certain additional restrictions, the fact that $n^{1/2}X_n$ is asymptotically normal with mean 0 and a certain variance. It is easy to check that this result can be obtained, under Assumptions (B1) through (B6) and the assumption that $M(x - c_0) - M(x + c_0)$ is differentiable at $x = \theta$, by using Theorem 1, replacing the $M(x)$ in Theorem 1 by $[M(x - c_0) - M(x + c_0)] / c_0$. Since this modification depends on knowing c_0 it will usually be undesirable. Theorem 2 (also 2′) gives a result using the Kiefer-Wolfowitz procedure which has the advantage of not depending on c_0. This gain, however, is offset, if c_0 is known, by the fact that, in general, for the Kiefer-Wolfowitz procedure X_n can never be $O_p(n^{-1/2})$ (see Example 1). However, as noted in the remarks following the proof of Theorem 2, $\{a_n\}$ and $\{c_n\}$ can be chosen so that X_n is arbitrarily close to being $O_p(n^{-1/2})$ without ever attaining it.

Under a stronger set of Assumptions than (B1) through (B6) Derman [6] proves a weaker result than the one we prove in Theorem 2. Using Chung's methods he shows that for any $t < 1/2$ there exist sequences $\{a_n\}$ and $\{c_n\}$ such that $n^t(X_n - \theta)$ is asymptotically normal with mean 0 and a certain variance.

THEOREM 2. *Suppose that Assumptions* (B1) *through* (B6) *are satisfied. Let* $AK_1 > 1/2$ *and take*

$$a_n = An^{-1} \tag{4.10}$$

Let $\{c_n\}$ *be a sequence of positive numbers satisfying* (4.0) *with* $a_n = An^{-1}$, *and the assumptions of Lemma* 5 *with* $r = 0$. *Then* $n^{1/2}c_n(X_n - \theta)$ *is asymptotically normally distributed with mean* 0 *and variance* $\sigma^2A^2(8\alpha A - 1)^{-1}$.

PROOF. With no loss of generality we assume $\alpha_0 = \theta = 0$. Abbreviating $\delta(X_n - c_n) - \delta(X_n + c_n)$ by δ_n and using (B2), we rewrite (4.2) and obtain

$$X_{n+1} = (1 - 4\alpha An^{-1})X_n - An^{-1}c_n^{-1}\delta_n - An^{-1}c_n^{-1}Z_n \tag{4.11}$$

Let $a = 4\alpha A$. Using the notation of (2.1) with $a_j = aj^{-1}$, iteration of (4.11) yields

(4.12) $$X_{n+1} = \beta_{on} X_1 - A \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} \delta_m - A \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} Z_m$$

Let $h_n = (\sum_{m=1}^{n} a^2 m^{-2} c_m^{-2} \beta_{mn}^2)^{-1/2}$. By Lemma 5 with $r = 0$ and $\rho = 2$ we have

(4.13) $$h_n^2 \sim a^{-2}(2a - 1) n c_n^2$$

Hence, what we wish to prove is that

(4.14) $a(2a - 1)^{-1/2} h_n X_n$ is asymptotically normal with mean 0 and variance $A^2\sigma^2(8\alpha A - 1)^{-1}$

After multiplying both sides of (4.13) by $(2a - 1)^{-1/2} a h_n$ it becomes clear that (4.14) will be proved if we can prove

(4.15a) $$h_n \beta_{on} \rightarrow 0,$$

(4.15b) $$h_n \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} \delta_m \rightarrow 0 \quad \text{in probability,}$$

and

(4.15c) $h_n \sum_{m=1}^{n} a m^{-1} c_m^{-1} \beta_{mn} Z_m$ is asymptotically normal with mean 0 and variance σ^2.

Lemma 2 shows that (4 15a) holds. We establish (4 15c) by using the same argument used to prove (3 9c) in Theorem 1 The details being the same we omit the argument except to note that, by Lemma 3, and (B5),

$$\sigma_n^2 = h_n^2 \sum_{m=1}^{n} a^2 m^{-2} c_m^{-2} \beta_{mn}^2 E Z_m^2 \rightarrow \sigma^2$$

Note that up to this point the only assumptions used have been (B1), (B2), (B5), and (B6). This observation will enable us to begin the proofs of later theorems at the point where we have to verify (4.15b).

To establish (4.15b) we require an estimate of EX_n^2 which we now obtain.

Squaring both sides of (4.2), taking expected values, and making use of (B3) and (B5) we obtain, for $n > N_0$ where N_0 is large enough so that $c_{N_0} < c_0$,

(4.16) $$EX_{n+1}^2 \leqq (1 - 2AK_1 n^{-1} + A^2 K_2^2 n^{-2}) EX_n^2 + s A^2 c_n^{-2} n^{-2}$$

Let $u > 0$ be such that $2AK_1 - u > 1$ and denote $2AK_1 - u$ by p. Then, for sufficiently large n, say $n \geqq N_1$, (4.16) implies

(4.17) $$EX_{n+1}^2 \leqq (1 - p n^{-1}) EX_n^2 + D_1 c_n^{-2} n^{-2}$$

Put $a_j = pj^{-1}$ in (2.1) and denote the β_{mn} thus obtained by β'_{mn}. Then, iterating (4 17) and using (2.3) and Lemma 4 with $c_n^{-2} = d_n$ and $q = p - 2$, we obtain,

for $n > N_1$,

$$(4.18)\qquad \begin{aligned} EX_{n+1}^2 &\leqq \beta'_{N_1 n} EX_{N_1+1}^2 + D_1 \sum_{m=N_1+1}^{n} m^{-2} c_m^{-2} \beta'_{mn} \\ &\leqq D_2 n^{-p} + D_3 \sum_{m=N_1+1}^{n} m^{p-2} c_m^{-2} n^{-p} = O(c_n^{-2} n^{-1}) \end{aligned}$$

which is the desired estimate.

For each integer m define ϕ_m to be 1 if $|X_m| \leqq c_m$ and $\phi_m = 0$ for $|X_m| > c_m$. Then, to prove (4.15b) it is sufficient to prove

$$(4.19a)\qquad h_n \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} \delta_m \phi_m \to 0 \quad \text{in probability}$$

$$(4.19b)\qquad h_n \sum_{m=1}^{n} m^{-1} c_m^{-1} \beta_{mn} \delta_m (1 - \phi_m) \to 0 \quad \text{in probability}$$

For $\epsilon > 0$ it is a consequence of (B4), that, for m sufficiently large, say $m > N_2$, $|\phi_m \delta_m c_m^{-1}| \leqq \epsilon |X_m|$. Use of Lemma 2 and a Chebyshev-type inequality now show that (4.19a) is implied by

$$(4.20a)\qquad h_n \sum_{m=N_2}^{n} m^{-1} \beta_{mn} E^{1/2}(X_m^2) = O(1)$$

Since $\delta_m c_m^{-1} = O(|X_m|)$ (a consequence of (B3)) and

$$E(|X_m|(1 - \phi_m)) \leqq P^{1/2}\{|X_m| > c_m\} E^{1/2}(X_m^2)$$

it follows, in similar fashion, that (4.19b) is implied by

$$(4.20b)\qquad h_n \sum_{m=1}^{n} m^{-1} \beta_{mn} E^{1/2}(X_m^2) P^{1/2}\{|X_m| > c_m\} = o(1)$$

Since our assumptions on $\{c_n\}$ imply that $c_n n^{1/4} \to \infty$ (see the proof of Lemma 4 where it is shown that $d_n n^{q+1} \to \infty$ for $q + 1 > 0$) we have $P\{|X_m| > c_m\} \leqq c_m^{-2} EX_m^2 = 0(c_m^{-4} m^{-1}) \to 0$ as $m \to \infty$. Hence (4.20b) will follow from (4.20a) by an argument like that in Lemma 3.

To show (4.20a) observe that by (2.3), (4.18), and Lemma 4 with $d_m = c_m^{-1}$ and $q = a - 3/2$ ($a - 3/2 > -1$ because our assumptions imply that $4\alpha \geqq K_1$ and hence $a = 4\alpha A \geqq AK_1 > 1/2$) we have, for $n > N_3 = \max(N_1, N_2)$,

$$(4.21)\qquad h_n \sum_{N_3}^{n} m^{-1} \beta_{mn} E^{1/2}(X_m^2) \leqq D_4 h_n n^{-a} \sum_{N_3}^{n} m^{a-3/2} c_m^{-1} \leqq D_5 h_n n^{-1/2} c_n^{-1}.$$

Use of (4.13) and Lemma 2 yields (4.20a) thus completing the proof of Theorem 2.

THEOREM 2′. *Suppose that all the conditions of Theorem 2 are satisfied except that* (B3) *is replaced by* (B3′). *Then* $n^{1/2} c_n (X_n - \theta)$ *is asymptotically normal with mean* 0 *and variance* $\sigma^2 A^2 (8\alpha A - 1)^{-1}$.

We omit the proof since it follows from Theorem 2 by use of the same truncation argument used in obtaining Theorem 1′ from Theorem 1.

It is easily checked that, for any sequence $\{f_n\}$ of positive numbers approaching 0, there exists a sequence $\{c_n\}$ satisfying the conditions of Theorem 2 and such that $c_n \geqq f_n$. Thus Theorem 2 says that under its conditions we can always find sequences $\{a_n\}$ and $\{c_n\}$ satisfying (4.0) and such that X_n is arbitrarily close to being $O_p(n^{-1/2})$ without ever attaining it. The question then arises as to whether it is possible to choose $\{a_n\}$ and $\{c_n\}$ satisfying (4.0) and such that $X_n = O_p(n^{-1/2})$. The answer to this is, in general, negative. To see this we give the following example.

EXAMPLE 1. Let $M(x) = -x^2/4$. For each x let $Z(x)$ be normally distributed with mean 0 and variance 1/2, and let $Z(X_m - c_m)$ and $Z(X_m + c_m)$ be independent. Then $\{Z_m\}$ is a sequence of independent normal random variables with mean 0 and variance 1. Note also that $\{Z_m\}$ and X_1 are independent. Let $\{a_n\}$ and $\{c_n\}$ be sequences of positive numbers satisfying (4.0). We now show that it is impossible that, for any infinite sequence $\{n_k\}$ of distinct integers,

$$\lim_{y\to\infty} \lim_{k\to\infty} P\{n_k^{1/2} X_{n_k+1} < y\} = 1. \tag{4.22}$$

For, if it were possible, writing $X_{n+1} = \beta_{on} X_1 - \sum_{m=1}^{n} a_m c_m^{-1} \beta_{mn} Z_m$, we would have

$$\lim_{y\to\infty} \lim_{k\to\infty} P\left\{ n_k^{1/2} \sum_{m=1}^{n_k} a_m\, c_m^{-1} \beta_{mn_k}\, Z_m < y \right\} = 1$$

which, by the normality of Z_m, implies that

$$n_k \sum_{m=1}^{n_k} a_m^2\, c_m^{-2} \beta_{mn_k}^2 = O(1).$$

But this is impossible by Lemma 1 and the fact that

$$\left(\sum_{m=1}^{n_k} a_m\, c_m^{-1} \beta_{mn_k} \right)^2 \leqq n_k \sum_{m=1}^{n_k} a_m^2\, c_m^{-2} \beta_{mn_k}^2$$

For Theorem 3 we drop (B3) and (B4) and substitute in their place

ASSUMPTION (B7). There exist positive numbers ϵ, c_0, and K_1 with $\epsilon > c_0$ such that, for all $c \leqq c_0$ and all x satisfying $c < |x - \theta| < \epsilon$

$$(x - \theta)[M(x - c) - M(x + c)]c^{-1} > K_1(x - \theta)^2 \tag{4.23}$$

(B2) and (B7) are both implied by the condition (which we refer to hereafter as the derivative condition) that M has a continuous second derivative in a neighborhood of θ with $M''(\theta) = -2\alpha$. Under (B1), the derivative condition, (B5), and the assumption that (3.6) hold for *all* $v > 0$ and some $\epsilon > 0$, Burkholder produces, for every $t < 1/4$, sequences $\{a_n\}$ and $\{c_n\}$ for which $n^t(X_n - \theta)$ is asymptotically normal with mean 0 and a certain variance. Theorem 3 shows that under weaker restrictions the same is true for $t = 1/4$.

THEOREM 3. *Suppose that Assumptions* (B1), (B2), (B5), (B6) *and* (B7) *are satisfied with* $K_1 \leqq 4\alpha$ *in* (B7). *Let* $c_n = n^{-1/4}$ *and* $a_n = An^{-1}$ *where* A *is such that*

$AK_1 > 1/4$. *Then* $n^{1/4}(X_n - \theta)$ *is asymptotically normal with mean* 0 *and variance* $\sigma^2 A^2 (8\alpha A - 1)^{-1}$.

PROOF. Let $\alpha_0 = \theta = 0$. If we prove Theorem 3 when (B7) is strengthened so that (4.23) holds for all x satisfying $|x| > \epsilon$ and, in addition, $|M(x - c) - M(x + c)| \leqq K_2 |x|$ for all $|x| > \epsilon$ and all $c \leqq c_0$ then, by using the truncation device used in the proof of Theorem 1', we will be able to establish Theorem 3 with (B7) as it stands. By the remarks made in the proof of Theorem 2 we will be finished with this proof if we can verify (4.15b).

As previously we require an estimate of EX_n^2 obtained as follows. Let $\phi_n = 1$ if $|X_n| \leqq c_n$ and $\phi_n = 0$ if $|X_n| > c_n$. Let $t > 0$. Then, it is a consequence of (B2) that, for n sufficiently large, say $n > N_1$

$$|\delta_n| \phi_n \leqq tc_n^2 \tag{4.24}$$

Squaring (4.2), taking expected values, and using (B5), (B2), and the strengthened form of (B7) yields

$$\begin{aligned} EX_{n+1}^2 &\leqq E\phi_n(X_n - An^{-1}c_n^{-1}M_n)^2 + E(1 - \phi_n)(X_n - An^{-1}c_n^{-1}M_n)^2 \\ &\quad + D_8 n^{-2}c_n^{-2} \\ &\leqq E\phi_n(1 - 4\alpha A n^{-1})^2 X_n^2 + D_9 E\phi_n |\delta_n X_n| n^{-1}c_n^{-1} \\ &\quad + D_{10} E\phi_n \delta_n^2 n^{-2} c_n^{-2} \\ &\quad + E(1 - \phi_n)(1 - 2K_1 A n^{-1} + A^2 K_2^2 n^{-2} c_n^{-2}) X_n^2 + D_8 n^{-2} c_n^{-2} \end{aligned} \tag{4.25}$$

Let u and w be positive numbers such that $2K_1A - w > 1/2$ and $2K_1A - w < 8\alpha A - u$, and let $2K_1A - w$ be denoted by p. Choose $N_2 > N_1$ so that, if $n > N_2$, $A^2K_2^2 n^{-2}c_n^{-2} < wn^{-1}$ and $16\alpha^2A^2n^{-2} < un^{-1}$. Then, for all $n > N_2$, we have from (4.24) and (4.25)

$$\begin{aligned} EX_{n+1}^2 &\leqq (1 - pn^{-1})EX_n^2 + D_{11}n^{-2}c_n^{-2} + tD_9 n^{-1}c_n^2 \\ &= (1 - pn^{-1})EX_n^2 + D_{12}n^{-3/2} \end{aligned} \tag{4.26}$$

Iteration of (4.26) now shows that, for $n > N_2$,

$$EX_{n+1}^2 = O(n^{-1/2}) \tag{4.27}$$

which is the desired estimate.

(B2) and the fact that $X_n \to 0$ w.p.1 imply that

$$\lim_{n\to\infty} \delta_n (X_n^2 + c_n^2)^{-1} = 0$$

w.p.1. Hence an argument like that in Theorem 1((3.20) et seq) shows that in order to verify (4.15b) it is sufficient to prove that, for any integer $N > N_2$

$$h_n \sum_{m=N}^{n} m^{-1}c_m^{-1}\beta_{mn}(EX_m^2 + c_m^2) = O(1). \tag{4.28}$$

Putting $c_m = m^{-1/4}$ in (4.28) and using (4.27) this follows quite easily, thus finishing the proof of Theorem 3.

If we make some further assumptions about M we will be able to improve on the $n^{1/4}$ obtained in Theorem 3. To this end note that from (B2) we have

$$|\delta(x)| \leqq \epsilon_x(x-\theta)^2 \text{ where } \epsilon_x \to 0 \text{ as } x \to \theta.$$

Assumption (B8) which we now specify is an assumption about ϵ_x.

ASSUMPTION (B8). There exist positive numbers c_0, ρ, and R such that, for all $c \leqq c_0$,

$$\sup_{|x-\theta|\leqq c} \epsilon_x \leqq Rc^\rho$$

If $\delta(x, \theta) = O(|x-\theta|^3)$ for x near θ it is easy to see that (B8) is satisfied for appropriate R and c_0 and $\rho = 1$; thus, the case of most interest is when $\rho = 1$. (B8) is very closely related to Burkholder's condition of "local-evenness"—for $\rho \geqq 0$, M is called ρ-locally-even if

$$\limsup_{\epsilon\to 0} f(\epsilon)\epsilon^{-1-\rho} < \infty$$

where $f(\epsilon) = \sup\{x \mid M(x-\epsilon) - M(x+\epsilon) \leqq 0\}$. It is easy to verify that when M is continuous in a neighborhood of θ and (B8) is satisfied then M is ρ-locally-even. In fact, when M satisfies the derivative condition, requiring M to be ρ-locally-even is equivalent to requiring $\delta(x, \theta) - \delta(-x, \theta) = O(|x-\theta|^{2+\rho})$ as $x - \theta \to 0$. The disadvantage in assuming the slightly more restrictive (B8) rather than local-evenness is allayed by the fact that (B8) appears as a more natural condition.

Under (B1), the derivative condition, (B5), the assumption that (3 6) hold for all $v > 0$ and some $\epsilon > 0$, and the assumption that M is ρ-locally-even, Burkholder proves that, for any $t < (1+\rho)/(4+2\rho)$, there exist sequences $\{a_n\}$ and $\{c_n\}$ such that $n^t(X_n - \theta)$ is asymptotically normal. Theorem 4 replaces the condition of local-evenness by (B8), weakens the other assumptions made by Burkholder, and gives a stronger conclusion, e.g., with $\rho = 1$ in (B8), $a_n = An^{-1}$, and $c_n = (n^{1/6}\log n)^{-1}$, Theorem 4 shows that $n^{1/3}(\log n)^{-1}(X_n - \theta)$ is asymptotically normal.

THEOREM 4. *Suppose that Assumptions* (B1), (B2), (B5), (B6), (B7), *and* (B8) *are satisfied with* $K_1 \leqq 4\alpha$ *in* (B7). *Let* $a_n = An^{-1}$ *where* A *is such that* $AK_1 > 1$ *and let* $c_n = d_n n^{-r}$ *with* $d_n \to 0$ *and satisfying* (2.5) *of Lemma* 4 *and with* $r = (4+2\rho)^{-1}$. *Then* $n^{1/2}c_n(X_n - \theta)$ *is asymptotically normal with mean* 0 *and variance* $\sigma^2A^2(8\alpha A - 1)^{-1}$.

PROOF. Let $\alpha_0 = \theta = 0$. By the reasoning in the first paragraph of the proof of Theorem 3 we have only to verify (4.15b). To do so we require an estimate of EX_n^2 which is obtained in much the same way as (4.26) is obtained. In fact, using (B8) to replace (4.24) by

$$|\phi_n\delta_n| = O(c_n^{2+\rho}), \tag{4.29}$$

a repetition of the argument given in Theorem 3 shows that

$$EX_{n+1}^2 \leqq (1 - pn^{-1})EX_n^2 + D_{13}n^{-1}c_n^{2+\rho} + D_{14}n^{-2}c_n^{-2} \tag{4.30}$$

Iterating (4.30) and applying Lemma 4 and (2.3)—note that $p = 2K_1A - w > 1$ and that $p > r(2 + \rho)$—yields, for $n > N$ where N is chosen sufficiently large (how large can be determined by inspecting the proof of Theorem 3),

$$\begin{aligned}(4.31)\qquad EX_{n+1}^2 &\leqq O(n^{-p}) + D_{13}\sum_{m=N}^{n} m^{-1}c_m^{2+\rho}\beta'_{mn} + D_{14}\sum_{m=N}^{n} m^{-2}c_m^{-2}\beta'_{mn}\\ &= 0(n^{-p}) + O(c_n^{2+\rho}) + O(n^{-1}c_n^{-2})\\ &= O(c_n^{2+\rho}) + O(n^{-1}c_n^{-2})\end{aligned}$$

which is the desired estimate.

To prove (4.15b), it is sufficient to prove that for N_1 sufficiently large

$$(4.32)\qquad h_n \sum_{m=N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}\,\phi_m\,\delta_m \to 0 \quad \text{in probability}$$

$$(4.33)\qquad h_n \sum_{m=N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}(1 - \phi_m)\delta_m \to 0 \quad \text{in probability.}$$

By (4.29), (2.7) of Lemma 5, and Lemma 4—note that $4\alpha A > r(1 + \rho)$—we obtain, for $n > N_1$,

$$(4.34)\qquad \left| h_n \sum_{N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}\,\phi_m\,\delta_m \right| = O\left(h_n \sum_{N_1}^{n} m^{-1}c_m^{1+\rho}\beta_{mn}\right) = O(n^{1/2}c_n^{2+\rho})$$

which, by the choice of $\{c_n\}$, proves (4.32).

To show (4.33) we proceed as follows. Let $\mu_m = 1$ if $|X_m| \leqq c_0$ (c_0 here is the same as in (B8)) and let $\mu_m = 0$ otherwise. Using (B8) and, from (B2), the fact that $|\delta_m| = O(|X_m|)$ if $|X_m| > c_0$, we then have, for N_1 sufficiently large so that in particular $c_m < c_0$ for $m \geqq N_1$

$$\begin{aligned}&h_n \sum_{N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}(1 - \phi_m)\,|\delta_m|\\ (4.35)\qquad &= h_n \sum_{N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}(1 - \phi_m)\mu_m\,|\delta_m|\\ &\qquad + h_n \sum_{N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}(1 - \phi_m)(1 - \mu_m)\,|\delta_m|\\ &= O\left(h_n \sum_{N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}X_m^2\right) + O\left(h_n \sum_{N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}(1 - \mu_m)\,|X_m|\right)\end{aligned}$$

Now

$$h_n \sum_{N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}E\{(1 - \mu_m)\,|X_m|\} \leqq h_n c_0^{-1} \sum_{N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}EX_m^2$$

and since, by (4.31) and Lemma 4 (note that $4\alpha A > 1$),

$$h_n \sum_{N_1}^{n} m^{-1}c_m^{-1}\beta_{mn}EX_m^2 = O(n^{1/2}c_n^{2+\rho}) + O(n^{-1/2}c_n^{-2}),$$

our choice of c_n shows that the right hand side of (4.35) goes to 0 in probability. This establishes (4.33) and finishes the proof of the theorem.

Focusing our attention for the present on the case $\rho = 1$ (this is by no means necessary since all ensuing remarks can be suited to the cases where $\rho \neq 1$), we can ask whether or not it is possible to find sequences $\{a_n\}$ and $\{c_n\}$ satisfying (4.0) and a sequence $\{g_n\}$ such that, under Assumptions (B1), (B2), and (B5) to (B8), $g_n X_{n+1}$ is asymptotically normal and $g_n^{-1} = O(n^{-1/3})$. Example 2, which we now give, shows that the answer to this question is no.

EXAMPLE 2. Let $\{a_n\}$ and $\{c_n\}$ be sequences satisfying (4.0). For $0 < C < 1/6$ let $M(x)$ be defined as follows.

$$\begin{aligned} M(x) &= -x^2/4 + x^3 \quad \text{if} \quad |x| \leqq C \\ &= -x^2/4 + C^3 \quad \text{if} \quad x > C \\ &= -x^2/4 - C^3 \quad \text{if} \quad x < -C \end{aligned} \tag{4.36}$$

For each x let $Z(x)$ be normally distributed with mean 0 and variance 1/2 and let $Z(X_m - c_m)$ and $Z(X_m + c_m)$ be independently distributed. Thus $\{Z_m\}$ is a sequence of independent normal random variables with mean 0 and variance 1 and Z_m and $X_{m'}$ are independent if $m \geqq m'$. It is clear that (B1), (B2), (B5), (B6), (B7), and (B8) with $\rho = 1$ are all satisfied. Suppose that $\{g_n\}$ is a sequence of real numbers such that $g_n X_{n+1}$ converges in distribution to the normal distribution with mean 0 and variance v with $v \geqq 0$. Since $|g_n| X_{n+1}$ is then also asymptotically normal with mean 0 and variance v we can assume to begin with that, for all n, $g_n \geqq 0$. We will show that $\lim \sup_{n\to\infty} n^{-1/3} g_n = 0$.

Let $\phi_{1m}, \phi_{2m}, \cdots, \phi_{5m}$ be random variables taking on the values 0 and 1 only, with the value 1 being taken on as follows:

$$\begin{aligned} \phi_{1m} &= 1 \quad \text{if} \quad |X_m - c_m| \leqq C, |X_m + c_m| \leqq C \\ \phi_{2m} &= 1 \quad \text{if} \quad X_m - c_m > C \\ \phi_{3m} &= 1 \quad \text{if} \quad X_m - c_m \leqq C < X_m + c_m \\ \phi_{4m} &= 1 \quad \text{if} \quad X_m + c_m < -C \\ \phi_{5m} &= 1 \quad \text{if} \quad X_m - c_m < -C \leqq X_m + c_m \end{aligned}$$

Let N_0 be such that, for all $m > N_0$, $c_m < C/2$ and, in addition, suppose that N_0 is large enough so that $a_m < 1$ for all $m > N_0$—the latter requirement is to guarantee that $\beta_{mn} > 0$ for all $n \geqq m > N_0$. Since for all $m > N_0$, $\sum_{i=1}^{5} \phi_{im} = 1$, it follows from (4.36) that $m > N_0$ implies

$$\begin{aligned} M_m = M(X_m - c_m) - M(X_m + c_m) &= \sum_{i=1}^{5} M_m \phi_{im} \\ &= c_m X_m - 2c_m^3 \phi_{1m} - 6c_m X_m^2 \phi_{1m} + ((X_m - c_m)^3 - C^3)\phi_{3m} \\ &\quad - (C^3 + (X_m + c_m)^3)\phi_{5m}. \end{aligned}$$

Observe that none of the last three terms is positive. Abbreviating $-a_m/c_m$ times their sum by G_m we obtain from (4.2)

$$X_{n+1} = (1 - a_m)X_m + 2c_m^2 a_m \phi_{1m} + G_m - a_m c_m^{-1} Z_m \tag{4.37}$$

Iterating (4.37) we obtain, for $n > N \geqq N_0$,

$$\begin{aligned} X_{n+1} &= \beta_{Nn} X_{N+1} + 2 \sum_{N+1}^{n} a_m c_m^2 \beta_{mn} \phi_{1m} + \sum_{N+1}^{n} \beta_{mn} G_m - \sum_{N+1}^{n} a_m c_m^{-1} \beta_{mn} Z_m \\ &= \beta_{Nn} X_{N+1} + G_{1nN} + G_{2nN} + G_{3nN} \end{aligned} \tag{4.38}$$

where G_{1nN}, G_{2nN}, and G_{3nN} are abbreviations for the terms in the corresponding positions in the previous line and we note that G_{2nN} is never negative and that G_{3nN} is normally distributed with mean 0 and variance $\sum_{N+1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2$.

We will show that $\lim \sup_{n\to\infty} n^{-1/3} g_n = 0$ by contradicting the assumption that there exists a positive constant D_{15} and a subsequence $\{n_k\}$ such that $n_k^{-1/3} g_{n_k} \geqq D_{15}$ for all k. We may assume that $\{n_k\}$ consists of all the positive integers since the argument below remains valid if we begin by restricting ourselves to the subsequence $\{n_k\}$ for which $n_k^{-1/3} g_{n_k} \geqq D_{15}$. Let

$$H_{nN} = \left(\sum_{N+1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2\right)^{-1/2} \text{ and } G'_{nN} = 2 \sum_{N+1}^{n} a_m c_m^2 \beta_{mn}.$$

We will arrive at the contradiction by showing first that the asymptotic normality of $g_n X_{n+1}$ implies that $g_n H_{nN_0}^{-1} = O(1)$ as $n \to \infty$ and $\lim \sup_{n\to\infty} g_n G'_{nN} = o(1)$ as $N \to \infty$, and then showing (see (4.45) et seq) the impossibility of having simultaneously $H_{nN_0}^{-1} = O(n^{-1/3})$ and $\lim \sup_{n\to\infty} n^{1/3} G'_{nN} = o(1)$ as $N \to \infty$.

Let E_m be the set $\{ | X_j | \leqq C/2, j > m\}$. Since $X_m \to 0$ w.p.1, $1 - P\{E_m\} = \epsilon_m \to 0$ as $m \to \infty$. Since $c_m < C/2$ for all $m > N_0$ we have, for all such m,

$$E_m \subset \{\phi_{1j} = 1, j > m\}. \tag{4.39}$$

For $v \geqq 0$ let F_v denote the normal distribution with mean 0 and variance v. We consider two cases according as $v = 0$ or $v > 0$.

Case 1: $v = 0$. To begin with we obtain from (4.38), the fact that $G_{1nN} + G_{2nN}$ is never negative, and the independence of X_{N+1} and G_{2nN} that, for all $n > N_0$,

$$\begin{aligned} P\{X_{n+1} > 0\} &= P\{\beta_{N_0 n} X_{N_0+1} + G_{1nN_0} + G_{2nN_0} + G_{3nN_0} > 0\} \\ &\geqq P\{\beta_{N_0 n} X_{N_0+1} + G_{3nN_0} > 0\} \\ &\geqq P\{X_{N_0+1} > 0, G_{3nN_0} > 0\} \\ &= \tfrac{1}{2} P\{X_{N_0+1} > 0\} \end{aligned} \tag{4.40}$$

We will show that for some N $\lim_{n\to\infty} g_n G'_{nN} = 0$. Since G'_{nN} is decreasing in N this will imply $\lim_{N\to\infty} \lim \sup_{n\to\infty} g_n G'_{nN} = 0$.

Suppose that for all N $\lim \sup_{n\to\infty} g_n G'_{nN} > 0$. Then, for each N, there exists a positive constant D_{16} and a sequence $\{n_k\}$ such that, for all k, $g_{n_k} G'_{n_k N} > D_{16}$. Let $G^*_{nN} = g_n G_{3nN}$. Then, since $g_n X_{n+1} \to 0$ in probability we have, by (4.38),

(4.39) and (4.40),

$$0 = 1 - F_0(D_{16}) \geqq \lim_{k\to\infty} P\{g_{n_k}\beta_{Nn_k}X_{N+1} + G^*_{n_kN} + g_{n_k}G_{1n_kN} + g_{n_k}G_{2n_kN} > D_{16}, E_N\}$$

(4.41)
$$\geqq \lim_{k\to\infty} P\{g_{n_k}\beta_{Nn_k}X_{N+1} + G^*_{n_kN} > 0; E_N\}$$

$$\geqq \lim_{k\to\infty} P\{X_{N+1} > 0, G^*_{n_kN} > 0\} - \epsilon_N$$

$$\geqq \tfrac{1}{4}P\{X_{N_0+1} > 0\} - \epsilon_N$$

Since N can be chosen large enough so that the right-hand side of (4.41) is strictly positive we have a contradiction, thus proving that $g_nG'_{nN} = o(1)$ for some N.

To show that $g_nH^{-1}_{nN_0} = O(1)$ assume, to the contrary, that $g_{n_k}H^{-1}_{n_kN_0} \to \infty$ for some sequence $\{n_k\}$. Since $H_{nN_0}\,G_{3nN_0}$ is normally distributed with mean 0 and variance 1 we would have, for any y,

$$\lim_{k\to\infty} P\{G^*_{n_kN_0} > y\} = \tfrac{1}{2}$$

Hence

$$0 = 1 - F_0(y) \geqq \lim_{k\to\infty} P\{g_{n_k}\beta_{N_0n_k}X_{N_0+1} + G^*_{n_kN_0} > y\}$$

$$\geqq \lim_{k\to\infty} P\{G^*_{n_kN_0} > y\}P\{X_{N_0+1} > 0\}$$

$$\geqq \tfrac{1}{2}P\{X_{N_0+1} > 0\}$$

which is a contradiction, thus proving that $g_nH^{-1}_{nN_0} = O(1)$.

Case 2: $v > 0$. The argument used in Case 1 to show that $g_nH^{-1}_{nN_0} = O(1)$ can be used here with (4.42) becoming

$$1 - F_v(y) \geqq \tfrac{1}{2}P\{X_{N_0+1} > 0\}$$

Let $y \to \infty$ and we obtain a contradiction to the assumption that $g_nH^{-1}_{nN_0}$ is not $O(1)$.

To show that $\limsup_{n\to\infty} g_n\,G'_{nN} = o(1)$ suppose, to the contrary, there exists a sequence $\{N_j\}$, a positive number D_{17}, and a sequence $\{n_{k_j}\}$ such that, for all j, k, $g_{n_{k_j}}\,G'_{n_{k_j}N_j} > D_{17}$. Let T be a random variable independent of

$$\{G_{3nN}, n > N, N > 0\}$$

and having F_v as its distribution function. Then

(4.43)
$$\sup_y |P\{g_nX_{n+1} < y\} - P\{T < y\}| < t_n$$

where $t_n \to 0$ as $n \to \infty$. Then, letting $u_{kj} = g_{n_kj} \beta_{N_j n_{kj}} g_{N_j}^{-1}$, we have

$$
(4.44)\quad \begin{aligned}
1 - F_v(D_{17}) &\geqq \limsup_{k\to\infty} P\{u_{kj} g_{N_j} X_{N_j+1} + G^*_{n_kjN_j} > 0; E_{N_j}\} \\
&\geqq \limsup_{k\to\infty} P\{u_{kj} g_{N_j} X_{N_j+1} + G^*_{n_kjN_j} > 0\} - \epsilon_{N_j} \\
&\geqq \limsup_{k\to\infty} P\{u_{kj} T + G^*_{n_kjN_j} > 0\} - \epsilon_{N_j} - t_{N_j} \\
&= \tfrac{1}{2} - \epsilon_{N_j} - t_{N_j}
\end{aligned}
$$

For j large enough it is clear that $\epsilon_{N_j} + t_{N_j} < F_v(D_{17}) - \frac{1}{2}$ which gives the desired contradiction.

To conclude the argument we have to show that it is impossible to have $\lim_{N\to\infty} \limsup_{n\to\infty} n^{1/3} G'_{nN} = 0$ and $n^{1/3} H^{-1}_{nN_0} = O(1)$. If it were possible we would have, for $N \geqq N_0$,

$$
(4.45)\quad \begin{aligned}
\left(\sum_{m=N+1}^{n} a_m^{3/2}\beta_{mn}^{3/2}\right)^2 &\leqq \left(\sum_{m=N+1}^{n} a_m c_m^2 \beta_{mn}\right)\left(\sum_{m=N+1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2\right) \\
&\leqq \left(\sum_{N+1}^{n} a_m c_m^2 \beta_{mn}\right)\left(\sum_{N_0+1}^{n} a_m^2 c_m^{-2} \beta_{mn}\right) \\
&\leqq \epsilon_{nN} n^{-1}
\end{aligned}
$$

where $\lim_{N\to\infty} \limsup_{n\to\infty} \epsilon_{nN} = 0$. Hence, using Holder's inequality,

$$
(4.46)\quad \left(\sum_{N+1}^{n} a_m \beta_{mn}\right)^3 \leqq \left(\sum_{N+1}^{n} a_m^{3/2}\beta_{mn}^{3/2}\right)^2 (n - N) \leqq \epsilon_{nN}(n - N)n^{-1}
$$

Applying Lemma 1 with $W_m = W = 1$ we conclude that, for each $N > N_0$,

$$
1 \leqq \limsup_{n\to\infty} \epsilon_{nN}
$$

which is impossible.

The reason that we cannot have $n^{1/3}X_n$ asymptotically normal in Example 2 is clearly the upsetting character of G_{1nN}. The following example shows how we can obtain $n^{1/3}$, and even better, by considering the asymptotic behavior of $\{X_{n+1} - G'_{n0}\}$ instead of $\{X_n\}$. What this indicates is that the "bias" term, G'_{n0}, around which X_n becomes rapidly concentrated, is the dominant error. Of course, the improvement in the order of convergence is of little practical use since it is $X_n - \theta$ which matters.

Example 3. Let M be as in Example 2 and let $Z(x)$ satisfy (B5) and (B6). Note that M satisfies (B8) with $\rho = 1$. Let $a_n = An^{-1}$ for $A > 1$ and let $c_n = n^{-1/8} d$, where d_n satisfies the conditions of Lemma 4 and $d, \to 0$. We will show that $n^{1/2}c_n(X_{n+1} - G'_{n0})$ is asymptotically normal with mean 0 and variance $\sigma^2 A^2(2A - 3/4)^{-1}$. By (4.36) and the kind of argument used several times be-

fore we will succeed in doing so if we can show

$$h_n \sum_{N+1}^{n} a_m c_m^2 \beta_{mn}(1 - \phi_m) = o_p(1) \tag{4.41}$$

$$h_n \sum_{N+1}^{n} a_m \beta_{mn} X_m^2 = o_p(1) \tag{4.42}$$

where ϕ_m is 1 or 0 according as $|X_m| \leq C - c_m$ or not, and where $\beta_{mn} = \prod_{j=m+1}^{n} (1 - a_j)$. By use of Chebyshev's inequality (4.41) and (4.42) will be proved if we show

$$h_n \sum_{N+1}^{n} a_m \beta_{mn} E X_m^2 = o(1). \tag{4.43}$$

Using (4.31) and Lemma 4 (note that $A > 1$) we obtain

$$\begin{aligned} h_n \sum_{N+1}^{n} m^{-1} \beta_{mn} E X_m^2 &= O\left(h_n \sum_{1}^{n} m^{-1} c_m^3 \beta_{mn}\right) + O\left(h_n \sum_{1}^{n} m^{-2} c_m^{-2} \beta_{mn}\right) \\ &= O(n^{1/2} c_n^4) + O(n^{-1/2} c_n^{-1}) = o(1) \end{aligned} \tag{4.44}$$

which proves (4.43). Note that we can do better than $n^{1/3}$ and, in fact, we can get arbitrarily close to $n^{3/8}$.

Blum in [3] has suggested a procedure which replaces (4.1) by

$$X_{n+1} = X_n - a_n c_n^{-1}[Y(X_n) - Y(X_n + c_n)].$$

This was suggested mainly for the multi-dimensional case which we consider in the next section but we point out here, in Example 4, that this procedure can be rather inefficient.

EXAMPLE 4. Let $M(x) = -x^2/2$ and let Z be as in Example 1. Then, using the Blum procedure,

$$X_{n+1} = \beta_{0n} X_1 - \frac{1}{2} \sum_{1}^{n} a_m c_m \beta_{mn} - \sum_{1}^{n} a_m c_m^{-1} \beta_{mn} Z_m$$

We show that if $h_n X_n = O_p(1)$ then h_n^{-1} cannot be $o(n^{-1/4})$. Theorem 2 shows, of course, that, for the Kiefer-Wolfowitz procedure, h_n^{-1} can be almost $O(n^{-1/2})$

If $h_n^{-1} = o(n^{-1/4})$ and $X_1 < 0$ we would have

$$\sum_{1}^{n} a_m c_m \beta_{mn} = o(n^{-1/4})$$

$$\sum_{1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2 = o(n^{-1/2})$$

Hence

$$\sum_{1}^{n} a_m^{4/3} \beta_{mn}^{4/3} \leq \left(\sum_{1}^{n} a_m c_m \beta_{mn}\right)^{2/3} \left(\sum_{1}^{n} a_m^2 c_m^{-2} \beta_{mn}^2\right)^{1/3} = o(n^{-1/3})$$

But then

$$\sum_1^n a_m \beta_{mn} \leqq \left(\sum_1^n a_m^{4/3} \beta_{mn}^{4/3} \right)^{3/4} n^{1/4} = o(1)$$

which is impossible by Lemma 1.

Again it is the "bias" term $\sum a_m c_m \beta_{mn}$ which is the dominant error, i.e., the Blum procedure becomes rapidly concentrated about the wrong value just as in Example 3.

5. Multi-Dimensional Procedures. In this section we consider multi-dimensional analogues of the Robbins-Monro and Kiefer-Wolfowitz procedures. Since the theorems and proofs for the multi-dimensional case are quite similar to those for the one dimensional case considered in Sections 3 and 4 we will not go into great detail in this section. We first consider a q-dimensional analogue of the Robbins-Monro procedure identical with the one considered by Blum [3]. The q-dimensional analogue of the Kiefer-Wolfowitz procedure considered next differs somewhat from the procedure given by Blum—the differences are pointed out below. At the end of the section we remark on some more general q-dimensional analogues.

Let x be a q-vector and let M be a vector-valued function of x with $M(x)$ also being a q-vector. Let α be a vector and let θ be a solution of the equation $M(x) = \alpha$. Let $Y(x)$ be a vector random variable with $EY(x) = M(x)$. The Robbins-Monro procedure for "locating" θ is given as follows.

Let $\{a_n\}$ be a sequence of positive real numbers such that

$$\sum a_n = \infty, \qquad \sum a_n^2 < \infty \tag{5.0}$$

Let X_1 be an arbitrary vector (as in Section 3 X_1 can actually be taken to be a random variable) and define $\{X_n, n \geqq 2\}$ by the recursion

$$X_{n+1} = X_n - a_n(Y(X_n) - \alpha) \tag{5.1}$$

where $Y(X_n)$ is a random variable whose conditional distribution given $X_1 = x_1, \cdots, X_n = x_n$ is the same as $Y(x_n)$. Writing $Y(x) = M(x) + Z(x)$ we obtain from (5.1)

$$X_{n+1} = X_n - a_n(M(X_n) - \alpha) - a_n Z(X_n) \tag{5.2}$$

where, as before, the conditional distribution of $Z(X_n)$ given $X_1 = x_1, \cdots, X_n = x_n$ is the same as the distribution of $Z(x_n)$ and

$$E(Z(X_n) \mid X_1, \cdots, X_n) = 0 \tag{5.3}$$

w.p.1.

The assumptions we make now are easily seen to correspond to the assumptions made in Section 3—(A2*) corresponding, of course to (A2′). The notation we use is the same as that adopted in Section 2 for Lemma 6.

Assumption (A1*). M is Borel-measurable, $M(\theta) = \alpha$, and, for every $\epsilon > 0$

$$\inf_{1/\epsilon > |x-\theta| > \epsilon} [x - \theta, M(x) - \alpha] > 0$$

(A1*) is satisfied, for example, if (A3*) is satisfied with $\delta = 0$ which is, of course, much stronger than needed.

ASSUMPTION (A2*). There exists a positive constant K_1 such that, for all x,

$$| M(x) - \alpha | \leqq K_1 | x - \theta |$$

ASSUMPTION (A3*). For all x

$$M(x) = \alpha + B(x - \theta) + \delta(x, \theta)$$

where B is a positive definite $q \times q$ matrix and $| \delta(x, \theta) | = o(| x - \theta |)$ as $x - \theta \to 0$.

ASSUMPTION (A4*).

$$\sup_x E | Z(x) |^2 < \infty \tag{5.4}$$

$$\lim_{x \to \theta} EZ(x)Z'(x) = \pi \tag{5.5}$$

where π is a non-negative definite matrix and where the limit is in the sense of the norm we have defined.

ASSUMPTION (A5*).

$$\lim_{R \to \infty} \lim_{\epsilon \to 0^+} \sup_{|x - \theta| < \epsilon} \int_{\{|Z(x)| > R\}} | Z(x) |^2 \, dP = 0$$

The remarks concerning Assumption (A5) in Section 3 also pertain here—we use (A5*) in conjunction with the convergence of X_n to θ w.p.1 (a consequence of (A1*), (A2*) and (5.4)) and (5.4) only to obtain

$$\lim_{R \to \infty} \sup_k \int_{\{|Z(X_k)| > R\}} | Z(X_k) |^2 \, dP = 0 \tag{5.6}$$

As before, with $Z(x)$ considered as a vector, (3.5) and (3.6) imply (A5*).

Let $b_1, \cdots, b_q$ denote the eigenvalues of B in decreasing order. Write $B = PDP^{-1}$ where P is orthogonal and D is the diagonal matrix whose diagonal elements are $b_1, \cdots, b_q$. Observe that $\inf_{|x|=1} [Bx, x] = b_q$, $\inf_{|x|=1} [Bx, Bx = b_q^2$, and $\| B \| = b_1$. Let π_{ij} be the (i, j)th element of π and let π_{ij}^* be the (i, j)th element of $\pi^* = P^{-1}\pi P$.

THEOREM 5. *Suppose that Assumptions* (A1*) *through* (A5*) *are satisfied. Let* $a_n = An^{-1}$ *where* A *is such that* $Ab_q > \frac{1}{2}$. *Then* $n^{1/2}(X_n - \theta)$ *is asymptotically normal with mean* 0 *and covariance matrix* PQP^{-1} *where* Q *is the matrix whose* (i, j)th *element is* $A^2(Ab_i + Ab_j - 1)^{-1} \pi_{ij}^*$.

PROOF. Let $\alpha = \theta = 0$. Let $u = P^{-1}x$, $M^*(u) = P^{-1}M(Pu)$, $\delta^*(u) = P^{-1}\delta(Pu)$, and $Z^*(u) = P^{-1}Z(Pu)$. Then, with x, M, δ, and Z being replaced by u, M^*, δ^*, and Z^* respectively, it is easy to see that (A1*) through (A5*) are satisfied with B replaced by D and π replaced by π^* and that (5.2) is transformed into another Robbins-Monro procedure with α replaced by $P^{-1}\alpha$. Thus, in order to prove the theorem it is sufficient to prove that, when B is diagonal, $n^{1/2}X_n$ is asymptotically normal with mean 0 and covariance matrix

$$((A^2(Ab_i + Ab_j - 1)^{-1}\pi_{ij})).$$

(A1*), (A2*), and (5.4) imply that X_n converges to 0 w.p.1 (this follows from Dvoretzky's theorem—Blum's earlier proof of convergence w.p.1 is under stronger assumptions) and, hence, using (A3*), an argument like that in Theorem 1' shows that we can add the additional restriction that there exists a positive constant K such that $AK > 1$, $K < b_q$, and, for all x,

$$[M(x), x] \geqq K \mid x \mid^2 \tag{5.7}$$

The proof proceeds now just as in Theorem 1. Iterating (5.2) and using (A3*) we obtain

$$X_{n+1} = B_{0n} X_1 - A \sum_{m=1}^{n} m^{-1} B_{mn} \delta_m - A \sum_{m=1}^{n} m^{-1} B_{mn} Z_m \tag{5.8}$$

where

$$B_{mn} = \prod_{m+1}^{n} (I - Aj^{-1}B). \quad \text{Let} \quad h_n = \left(\sum_{1}^{n} A^2 m^{-2} \| B_{mn} \|^2 \right)^{-1/2}.$$

Since $\| B_{mn} \| = (1 + \epsilon_m)(mn^{-1})^{Ab_q}$ where $\epsilon_m \to 0$ as $m \to \infty$, we have $h_n \sim (2Ab_q - 1)^{1/2} A^{-1} n^{1/2}$. Making use of (5.7) and the same argument as used to obtain (3.18) in Theorem 1 we obtain $E \mid X_m \mid^2 = O(m^{-1})$. It then follows just as in Theorem 1 that

$$h_n \left(B_{0n} X_1 - A \sum_{m=1}^{n} m^{-1} B_{mn} \delta_m \right) \to 0$$

in probability.

To conclude the proof we will apply Lemma 6 with $U_{nk} = Ah_n k^{-1} B_{kn} Z_k$. Just as in Theorem 1 we obtain quite readily that (2.19) and (2.20) are satisfied with this choice of U_{nk}. Since $\mid B_{kn} x \mid \geqq (1 + \epsilon_k)(kn^{-1})^{Ab_1} \mid x \mid$ it follows from (A5*) in the same way as in Theorem 1 that (2.21) is satisfied. We have only to compute $\lim_{n\to\infty} s_n$ to be finished. Let the (i, j)th elements of $EZ_k Z_k'$ and s_n be denoted by $\pi_{ij}^{(k)}$ and s_n^{ij} respectively. Let $\beta_{ikn} = \prod_{k+1}^{n} (1 - Ab_i j^{-1})$. Then

$$s_n^{ij} = A^2 h_n^2 \sum_{k=1}^{n} k^{-2} \beta_{ikn} \beta_{jkn} \pi_{ij}^{(k)}$$

Since $\pi_{ij}^{(k)} \to \pi_{ij}$ and $h_n^2 \sim (2Ab_q - 1) A^{-2} n$ it follows that

$$s_n^{ij} \to (2Ab_q - 1)(Ab_i + Ab_j - 1)^{-1} \pi_{ij}$$

Thus, when B is diagonal, $n^{1/2} X_n$ is asymptotically normal with mean 0 and covariance matrix $((A^2(Ab_i + Ab_j - 1)^{-1} \pi_{ij}))$, and this finishes the proof of the theorem.

We will now take up the multi-dimensional Kiefer-Wolfowitz procedure. Let x be a q-vector and let f be a real valued function of x. Let $y(x)$ be a real random variable with $Ey(x) = f(x)$. We will consider the following q-dimensional version of the Kiefer-Wolfowitz procedure for finding the point at which f has a maximum.

Let $\{a_n\}$ and $\{c_n\}$ be two sequences of positive real numbers satisfying

(5.9) $$\sum a_n = \infty, \qquad \sum a_n^2 c_n^{-2} < \infty, \qquad \lim_{n\to\infty} c_n = 0$$

For $1 \leqq i \leqq q$ let e_i be the q-vector whose ith coordinate is 1 and whose other coordinates are 0. Let $Y(x, a) = (y(x + ae_1), \cdots, y(x + ae_q))$. Let X_1 be an arbitrary q-vector and define $\{X_n, n \geqq 2\}$ by the recursion

(5.10) $$X_{n+1} = X_n - a_n c_n^{-1} (Y(X_n, -c_n) - Y(X_n, c_n))$$

where the conditional distribution of $Y(X_n, \pm c_n)$ given $X_1 = x_1, \cdots, X_n = x_n$ is the same as $Y(x_n, \pm c_n)$. Writing $y(x) = f(x) + z(x)$, and letting

$$M(x, a) = (f(x + ae_1), \cdots, f(x + ae_q)),$$

$$Z(x, a) = (z(x + ae_1), \cdots, (z(x + ae_q)),$$

we rewrite (5.10) and obtain

(5.11) $$X_{n+1} = X_n - a_n c_n^{-1}(M(X_n, -c_n) - M(X_n, c_n)) - a_n c_n^{-1}(Z(X_n, -c_n) - Z(X_n, c_n))$$

We will denote $M(X_n, -c_n) - M(X_n, c_n)$ by M_n and $Z(X_n, -c_n) - Z(X_n, c_n)$ by Z_n. It is clear that just as in Section 4

(5.12) $$E(Z_{n+1} \mid Z_1, \cdots, Z_n) = E(Z_{n+1} \mid X_1, \cdots, X_{n+1}) = 0$$

w.p.1.

The procedure we have defined by (5 10) differs from the one considered by Blum [3] in that Blum uses $Y(X_n, 0) - Y(X_n, c_n)$ rather than $Y(X_n, -c_n) - Y(X_n, c_n)$. The advantage of the Blum procedure is that it requires at each stage $q + 1$ observations whereas the number of observations required by (5.10) at each stage is $2q$. However, as noted in Example 4, the Blum procedure is quite inefficient with respect to the rate at which it converges to θ.

We now list the assumptions we require. The correspondence between these assumptions and those of Section 4 is easy to see.

ASSUMPTION (B1*). f is Borel-measurable, has a unique maximum at $x = \theta$, $|f(x + 1) - f(x)| \leqq D_1 + D_2|x|$ for some positive constants D_1 and D_2, and, for $0 < \epsilon_0 < \epsilon_1 < \epsilon_2 < \infty$,

$$\inf_{\substack{\epsilon_1 \leqq |x-\theta| \leqq \epsilon_2 \\ 0 < \epsilon \leqq \epsilon_0}} \epsilon^{-1}[M(x, -\epsilon) - M(x, \epsilon), x - \theta] > 0$$

(B1*) is satisfied, for example, if (B2*) is satisfied with $\delta = 0$; of course, this is much stronger than is needed.

ASSUMPTION (B2*). For all x

$$f(x) = \alpha_0 - [B(x - \theta), x - \theta] + \delta(x, \theta)$$

where α_0 is real, B is a positive definite $q \times q$ matrix, and $\delta(x, \theta) = o(|x - \theta|^2)$ as $x - \theta \to 0$.

(A1*), (A2*), and (5.4) imply that X_n converges to 0 w.p.1 (this follows from Dvoretzky's theorem—Blum's earlier proof of convergence w.p.1 is under stronger assumptions) and, hence, using (A3*), an argument like that in Theorem 1′ shows that we can add the additional restriction that there exists a positive constant K such that $AK > 1$, $K < b_q$, and, for all x,

$$[M(x), x] \geqq K \mid x \mid^2 \tag{5.7}$$

The proof proceeds now just as in Theorem 1. Iterating (5.2) and using (A3*) we obtain

$$X_{n+1} = B_{0n} X_1 - A \sum_{m=1}^{n} m^{-1} B_{mn} \delta_m - A \sum_{m=1}^{n} m^{-1} B_{mn} Z_m \tag{5.8}$$

where

$$B_{mn} = \prod_{m+1}^{n} (I - Aj^{-1}B). \quad \text{Let} \quad h_n = \left(\sum_{1}^{n} A^2 m^{-2} \| B_{mn} \|^2 \right)^{-1/2}.$$

Since $\| B_{mn} \| = (1 + \epsilon_m)(mn^{-1})^{Ab_q}$ where $\epsilon_m \to 0$ as $m \to \infty$, we have $h_n \sim (2Ab_q - 1)^{1/2} A^{-1} n^{1/2}$. Making use of (5.7) and the same argument as used to obtain (3.18) in Theorem 1 we obtain $E \mid X_m \mid^2 = O(m^{-1})$. It then follows just as in Theorem 1 that

$$h_n \left(B_{0n} X_1 - A \sum_{m=1}^{n} m^{-1} B_{mn} \delta_m \right) \to 0$$

in probability.

To conclude the proof we will apply Lemma 6 with $U_{nk} = Ah_n k^{-1} B_{kn} Z_k$. Just as in Theorem 1 we obtain quite readily that (2.19) and (2.20) are satisfied with this choice of U_{nk}. Since $\mid B_{kn} x \mid \geqq (1 + \epsilon_k)(kn^{-1})^{Ab_1} \mid x \mid$ it follows from (A5*) in the same way as in Theorem 1 that (2.21) is satisfied. We have only to compute $\lim_{n \to \infty} s_n$ to be finished. Let the (i, j)th elements of $EZ_k Z_k'$ and s_n be denoted by $\pi_{ij}^{(k)}$ and s_n^{ij} respectively. Let $\beta_{ikn} = \prod_{k+1}^{n} (1 - Ab_i j^{-1})$. Then

$$s_n^{ij} = A^2 h_n^2 \sum_{k=1}^{n} k^{-2} \beta_{ikn} \beta_{jkn} \pi_{ij}^{(k)}$$

Since $\pi_{ij}^{(k)} \to \pi_{ij}$ and $h_n^2 \sim (2Ab_q - 1) A^{-2} n$ it follows that

$$s_n^{ij} \to (2Ab_q - 1)(Ab_i + Ab_j - 1)^{-1} \pi_{ij}$$

Thus, when B is diagonal, $n^{1/2} X_n$ is asymptotically normal with mean 0 and covariance matrix $((A^2(Ab_i + Ab_j - 1)^{-1} \pi_{ij}))$, and this finishes the proof of the theorem.

We will now take up the multi-dimensional Kiefer-Wolfowitz procedure. Let x be a q-vector and let f be a real valued function of x. Let $y(x)$ be a real random variable with $Ey(x) = f(x)$. We will consider the following q-dimensional version of the Kiefer-Wolfowitz procedure for finding the point at which f has a maximum.

Let $\{a_n\}$ and $\{c_n\}$ be two sequences of positive real numbers satisfying

(5.9) $$\sum a_n = \infty, \qquad \sum a_n^2 c_n^{-2} < \infty, \qquad \lim_{n\to\infty} c_n = 0$$

For $1 \leqq i \leqq q$ let e_i be the q-vector whose ith coordinate is 1 and whose other coordinates are 0. Let $Y(x, a) = (y(x + ae_1), \cdots, y(x + ae_q))$. Let X_1 be an arbitrary q-vector and define $\{X_n, n \geqq 2\}$ by the recursion

(5.10) $$X_{n+1} = X_n - a_n c_n^{-1} (Y(X_n, -c_n) - Y(X_n, c_n))$$

where the conditional distribution of $Y(X_n, \pm c_n)$ given $X_1 = x_1, \cdots, X_n = x_n$ is the same as $Y(x_n, \pm c_n)$. Writing $y(x) = f(x) + z(x)$, and letting

$$M(x, a) = (f(x + ae_1), \cdots, f(x + ae_q)),$$

$$Z(x, a) = (z(x + ae_1), \cdots, (z(x + ae_q)),$$

we rewrite (5.10) and obtain

(5.11) $$X_{n+1} = X_n - a_n c_n^{-1}(M(X_n, -c_n) - M(X_n, c_n)) - a_n c_n^{-1}(Z(X_n, -c_n) - Z(X_n, c_n))$$

We will denote $M(X_n, -c_n) - M(X_n, c_n)$ by M_n and $Z(X_n, -c_n) - Z(X_n, c_n)$ by Z_n. It is clear that just as in Section 4

(5.12) $$E(Z_{n+1} \mid Z_1, \cdots, Z_n) = E(Z_{n+1} \mid X_1, \cdots, X_{n+1}) = 0$$

w.p.1.

The procedure we have defined by (5.10) differs from the one considered by Blum [3] in that Blum uses $Y(X_n, 0) - Y(X_n, c_n)$ rather than $Y(X_n, -c_n) - Y(X_n, c_n)$. The advantage of the Blum procedure is that it requires at each stage $q + 1$ observations whereas the number of observations required by (5.10) at each stage is $2q$. However, as noted in Example 4, the Blum procedure is quite inefficient with respect to the rate at which it converges to θ.

We now list the assumptions we require. The correspondence between these assumptions and those of Section 4 is easy to see.

Assumption (B1*). f is Borel-measurable, has a unique maximum at $x = \theta$, $|f(x + 1) - f(x)| \leqq D_1 + D_2 |x|$ for some positive constants D_1 and D_2, and, for $0 < \epsilon_0 < \epsilon_1 < \epsilon_2 < \infty$,

$$\inf_{\substack{\epsilon_1 \leqq |x-\theta| \leqq \epsilon_2 \\ 0<\epsilon\leqq\epsilon_0}} \epsilon^{-1}[M(x, -\epsilon) - M(x, \epsilon), x - \theta] > 0$$

(B1*) is satisfied, for example, if (B2*) is satisfied with $\delta = 0$; of course, this is much stronger than is needed.

Assumption (B2*). For all x

$$f(x) = \alpha_0 - [B(x - \theta), x - \theta] + \delta(x, \theta)$$

where α_0 is real, B is a positive definite $q \times q$ matrix, and $\delta(x, \theta) = o(|x - \theta|^2)$ as $x - \theta \to 0$.

ASSUMPTION (B3*). There exist positive numbers K_1, K_2, and c_0 such that, for all x in some neighborhood of θ and all c with $0 < c \leqq c_0$,

$$K_1 | x - \theta |^2 \leqq [x - \theta, (M(x, -c) - M(x, c))/c] \leqq K_2 | x - \theta |^2$$

and, for all x,

$$\left| \frac{M(x, -c) - M(x, c)}{c} \right| \leqq K_3 | x |.$$

ASSUMPTION (B4*). If $c_0 > 0$ then, for all x and c such that $| x | < c < c_0$

$$(\delta(x, -c, \theta) - \delta(x, +c, \theta))/c = o(| x - \theta |)$$

ASSUMPTION (B5*).

$$\sup_x E | Z(x, 0) |^2 < \infty \tag{5.13}$$

$$\lim_{\substack{x \to \theta \\ c \to 0}} E(Z(x, -c) - Z(x, c))(Z(x, -c) - Z(x, c))' = \pi \tag{5.14}$$

where π is a non-negative definite matrix.

ASSUMPTION (B6*).

$$\lim_{R \to \infty} \lim_{\epsilon \to 0^+} \sup_{|x| < \epsilon} \int_{\{|Z(x,0)| > R\}} | Z(x, 0) |^2 \, dP = 0$$

As before we use (B6*) to obtain

$$\lim_{R \to \infty} \sup_k \int_{\{|Z_k| > R\}} | Z_k |^2 \, dP = 0 \tag{5.15}$$

and, as before, (B6*) is implied by (3.5) or (3.6) with $Z(x)$ considered as a vector of course.

ASSUMPTION (B7*). There exist positive numbers ϵ, c_0, and K_1 such that, for all $c \leqq c_0$ and all x satisfying $c < | x - \theta | < \epsilon$,

$$[x - \theta, (M(x, -c) - M(x, c))/c] > K_1 | x - \theta |^2$$

Let $\delta(x, \theta) = \epsilon_x | x - \theta |^2$

ASSUMPTION (B8*). There exist positive numbers c_0, ρ, and R such that for all $c \leqq c_0$

$$\sup_{|x - \theta| \leqq c} \epsilon_x \leqq Rc^\rho$$

As in the paragraph preceding Theorem 5 let $B = PDP^{-1}$ and let $((\pi^*_{ij})) = \pi^* = P^{-1}\pi P$.

THEOREM 6. *Suppose Assumptions* (B1*) *through* (B6*) *are satisfied. Let* $AK_1 > 1/2$ *and choose* $a_n = An^{-1}$. *Let* $\{c_n\}$ *be a sequence of positive numbers satisfying* (5.9) *with* $a_n = An^{-1}$ *and the assumptions of Lemma* 5 *with* $r = 0$. *Then* $n^{1/2}c_n(X_n - \theta)$ *is asymptotically normal with mean* 0 *and covariance matrix* PQP^{-1} *where* $Q = ((A^2(4Ab_i + 4Ab_j - 1)^{-1} \pi^*_{ij}))$.

Choosing $t_1 = 1/b_1$ and $t_2 = 1/b_2$ will minimize the entries in the matrix in (5.21). Thus, if $b_1 \neq b_2$ we can do better by using $\{n^{-1}T\}$ than by using $\{An^{-1}\}$ since using $\{An^{-1}\}$ would correspond to the case where $t_1 = t_2$.

6. Concluding Remarks. In Sections 3, 4, and 5 we have restricted ourselves to sequences $\{a_n\}$ of the type $a_n = An^{-1}$. It is clear that arguments like the ones presented above can be given for cases where a_n is chosen to be something other than An^{-1} e.g., $a_n = An^{\epsilon-1}$. Due to Examples 1 and 2 however, the results of the previous sections are not likely to be improved very much by using these different sequences. Indeed, for the Robbins-Monro procedure it was shown in [5], Section 7 that under some restrictions, the Robbins-Monro procedure with $a_n = An^{-1}$ for a certain choice of A is optimal in the sense that it is asymptotically minimax for many weight functions. We may remark that this optimum property can be extended with no difficulty to the multi-dimensional Robbins-Monro procedure.

In [4] Burkholder considers somewhat more general processes than considered here in the sense that he permits $M(X_n)$ and $Z(X_n)$ to depend on n as well as X_n. With some modifications of the assumptions we have made this situation can be treated using the methods of Sections 3 and 4. Procedures given by Burkholder for locating points of inflection of a regression function and for finding the maximum of a density function can also be treated using our methods.

It is sometimes of interest to study the asymptotic behavior of $M(X_n) - \alpha$ for the Robbins-Monro procedure and of $M(X_n) - \alpha_0$ for the Kiefer-Wolfowitz procedure. It is easy to see that results about the asymptotic distribution of these quantities can be obtained from the results about the asymptotic distribution of $X_n - \theta$.

REFERENCES

[1] H. D. Block, "*On stochastic approximation*," ONR Report
[2] J. R. Blum, "Approximation methods which converge with probability one," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 382–386.
[3] J. R. Blum, "Multidimensional stochastic approximation procedures," *Ann Math. Stat.*, Vol. 25 (1954), pp. 737–744.
[4] D. L. Burkholder, "On a class of stochastic approximation procedures," *Ann Math Stat.*, Vol. 27 (1956), pp 1044–1059.
[5] K. L. Chung, "On a stochastic approximation method," *Ann Math Stat*, Vol. 25 (1954), pp 463–483.
[6] C Derman, "An application of Chung's lemma to the Kiefer-Wolfowitz stochastic approximation procedure," *Ann. Math. Stat*, Vol 27 (1956), pp 529–532
[7] A. Dvoretzky, "On stochastic approximation," *Proc of the Third Berkeley Symposium*, Vol. 1, pp. 39–56.
[8] J. L Hodges, Jr. and E. L Lehmann, "Two Approximations to the Robbins-Monro Process," *Proc. of the Third Berkeley Symposium*, vol 1, pp. 96–104.
[9] J Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat*, Vol 23 (1952), pp 462–466
[10] M. Loève, "*Probability theory*," New York (1955).
[11] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math Stat*, Vol. 22 (1951), pp. 400–407.
[12] J. Wolfowitz, "*On stochastic approximation methods*," *Ann Math Stat*, Vol 27 (1956), pp 1151–1155.

Assumption (B3*). There exist positive numbers K_1, K_2, and c_0 such that, for all x in some neighborhood of θ and all c with $0 < c \leqq c_0$,

$$K_1 | x - \theta |^2 \leqq [x - \theta, (M(x, -c) - M(x, c))/c] \leqq K_2 | x - \theta |^2$$

and, for all x,

$$\left| \frac{M(x, -c) - M(x, c)}{c} \right| \leqq K_3 | x |.$$

Assumption (B4*). If $c_0 > 0$ then, for all x and c such that $| x | < c < c_0$

$$(\delta(x, -c, \theta) - \delta(x, +c, \theta))/c = o(| x - \theta |)$$

Assumption (B5*).

(5.13) $$\sup_x E | Z(x, 0) |^2 < \infty$$

(5.14) $$\lim_{\substack{x \to \theta \\ c \to 0}} E(Z(x, -c) - Z(x, c))(Z(x, -c) - Z(x, c))' = \pi$$

where π is a non-negative definite matrix.

Assumption (B6*).

$$\lim_{R \to \infty} \lim_{\epsilon \to 0^+} \sup_{|x| < \epsilon} \int_{\{|Z(x,0)| > R\}} | Z(x, 0) |^2 \, dP = 0$$

As before we use (B6*) to obtain

(5.15) $$\lim_{R \to \infty} \sup_k \int_{\{|Z_k| > R\}} | Z_k |^2 \, dP = 0$$

and, as before, (B6*) is implied by (3.5) or (3.6) with $Z(x)$ considered as a vector of course.

Assumption (B7*). There exist positive numbers ϵ, c_0, and K_1 such that, for all $c \leqq c_0$ and all x satisfying $c < | x - \theta | < \epsilon$,

$$[x - \theta, (M(x, -c) - M(x, c))/c] > K_1 | x - \theta |^2$$

Let $\delta(x, \theta) = \epsilon_x | x - \theta |^2$

Assumption (B8*). There exist positive numbers c_0, ρ, and R such that for all $c \leqq c_0$

$$\sup_{|x - \theta| \leqq c} \epsilon_x \leqq R c^{\rho}$$

As in the paragraph preceding Theorem 5 let $B = PDP^{-1}$ and let $((\pi^*_{ij})) = \pi^* = P^{-1} \pi P$.

Theorem 6. *Suppose Assumptions* (B1*) *through* (B6*) *are satisfied. Let* $AK_1 > 1/2$ *and choose* $a_n = An^{-1}$. *Let* $\{c_n\}$ *be a sequence of positive numbers satisfying* (5.9) *with* $a_n = An^{-1}$ *and the assumptions of Lemma* 5 *with* $r = 0$. *Then* $n^{1/2} c_n (X_n - \theta)$ *is asymptotically normal with mean* 0 *and covariance matrix* PQP^{-1} *where* $Q = ((A^2 (4Ab_i + 4Ab_j - 1)^{-1} \pi^*_{ij}))$.

Choosing $t_1 = 1/b_1$ and $t_2 = 1/b_2$ will minimize the entries in the matrix in (5.21). Thus, if $b_1 \neq b_2$ we can do better by using $\{n^{-1}T\}$ than by using $\{An^{-1}\}$ since using $\{An^{-1}\}$ would correspond to the case where $t_1 = t_2$.

6. Concluding Remarks. In Sections 3, 4, and 5 we have restricted ourselves to sequences $\{a_n\}$ of the type $a_n = An^{-1}$. It is clear that arguments like the ones presented above can be given for cases where a_n is chosen to be something other than An^{-1} e.g., $a_n = An^{\epsilon-1}$. Due to Examples 1 and 2 however, the results of the previous sections are not likely to be improved very much by using these different sequences. Indeed, for the Robbins-Monro procedure it was shown in [5], Section 7 that under some restrictions, the Robbins-Monro procedure with $a_n = An^{-1}$ for a certain choice of A is optimal in the sense that it is asymptotically minimax for many weight functions. We may remark that this optimum property can be extended with no difficulty to the multi-dimensional Robbins-Monro procedure.

In [4] Burkholder considers somewhat more general processes than considered here in the sense that he permits $M(X_n)$ and $Z(X_n)$ to depend on n as well as X_n. With some modifications of the assumptions we have made this situation can be treated using the methods of Sections 3 and 4. Procedures given by Burkholder for locating points of inflection of a regression function and for finding the maximum of a density function can also be treated using our methods.

It is sometimes of interest to study the asymptotic behavior of $M(X_n) - \alpha$ for the Robbins-Monro procedure and of $M(X_n) - \alpha_0$ for the Kiefer-Wolfowitz procedure. It is easy to see that results about the asymptotic distribution of these quantities can be obtained from the results about the asymptotic distribution of $X_n - \theta$.

REFERENCES

[1] H. D. Block, "*On stochastic approximation*," ONR Report

[2] J. R. Blum, "Approximation methods which converge with probability one," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 382–386.

[3] J. R. Blum, "Multidimensional stochastic approximation procedures," *Ann Math Stat.*, Vol. 25 (1954), pp 737–744.

[4] D. L. Burkholder, "On a class of stochastic approximation procedures," *Ann Math Stat.*, Vol. 27 (1956), pp 1044–1059

[5] K. L. Chung, "On a stochastic approximation method," *Ann. Math. Stat*, Vol. 25 (1954), pp. 463–483.

[6] C. Derman, "An application of Chung's lemma to the Kiefer-Wolfowitz stochastic approximation procedure," *Ann. Math. Stat*, Vol. 27 (1956), pp. 529–532

[7] A. Dvoretzky, "On stochastic approximation," *Proc. of the Third Berkeley Symposium*, Vol. 1, pp. 39–56.

[8] J. L Hodges, Jr and E L Lehmann, "Two Approximations to the Robbins-Monro Process," *Proc. of the Third Berkeley Symposium*, vol 1, pp 95–104.

[9] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat*, Vol 23 (1952), pp 462–466.

[10] M. Loève, "*Probability theory*," New York (1955).

[11] H. Robbins and S. Monro, "A stochastic approximation method," *Ann Math. Stat*, Vol. 22 (1951), pp. 400–407.

[12] J. Wolfowitz, "On stochastic approximation methods," *Ann Math Stat*, Vol 27 (1956), pp. 1151–1155

THEOREM 7. *Suppose* (B1*), (B2*), (B5*), (B6*), *and* (B7*) *are satisfied wi* $K_1 \leqq 4b_q$ *in* (B7*). *Let* $c_n = n^{-1/4}$ *and* $a_n = An^{-1}$ *where* A *is such that* $AK_1 > 1/$ *Then* $n^{1/4}(X_n - \theta)$ *is asymptotically normal with mean* 0 *and covariance matr* PQP^{-1}.

We omit the proofs of Theorem 7 and Theorem 8 below since they procee from the proofs of Theorems 3 and 4 in the same fashion that the proof of Theo rem 6 did from that of Theorem 2.

THEOREM 8. *Suppose that* (B1*), (B2*), *and* (B5*) *through* (B8*) *are satisfi with* $K_1 \leqq 4b_q$ *in* (B7*). *Let* $a_n = An^{-1}$ *where* $AK_1 > 1$ *and let* $\{c_n\}$ *satisfy t conditions of Lemma* 5 *with* $d_n \to 0$ *and with* $r = (4 + 2\rho)^{-1}$. *Then* $n^{1/2}c_n(X_n -$ *is asymptotically normal with mean* 0 *and covariance matrix* PQP^{-1}.

The procedures given by (5.1) and (5.10) can be generalized if we replace $\{a$ by a sequence $\{T_n\}$ of matrices. When $\{T_n\}$ is a sequence of positive defini matrices such that, for all n, B and T_n are diagonalized by the same orthogon matrix P, and when the smallest and largest eigenvalues of T_n, denoted by and t_n^{**} respectively, satisfy (5.0) and (5.9) with a_n replaced by t_n^* and t_n^* methods like those used in the earlier part of this section and in earlier sectio can be used to study the asymptotic behavior of these procedures. Indeed, $T_n = n^{-1}T$ where T is a positive definite matrix which is diagonalized by results like those proved in the earlier part of this section can be obtained using the same methods as used in obtaining these results. When, for $A >$ $T = AI$, we are in the situation covered by those theorems. In studying (5.1 (the Kiefer-Wolfowitz procedure) we can, in addition, replace $\{c_n\}$ by a sequen $\{C_n\}$ of matrices; the remarks about $\{T_n\}$ are also relevant to $\{C_n\}$.

Since Examples 1 and 2 of Section 4 can be extended to their q-dimension analogues we cannot hope to improve materially the results of Theorems 5, 6, and 8 by using sequences $\{T_n\}$ which satisfy the second sentence of the precedi paragraph and which are more general than $\{a_nI\}$. However, if we knew B a π, then, by suitable choice of such $\{T_n\}$ we can, in general, obtain a limiti covariance matrix of smaller size than is obtainable by using merely $\{a_nI\}$. an indication of this suppose that we are concerned with a two-dimension Robbins-Monro procedure satisfying Assumptions (A1*) through (A5*) wi

$$\pi = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix}.$$

Letting

$$T_n = n^{-1}T = n^{-1}\begin{pmatrix} t_1 & 0 \\ 0 & t_2 \end{pmatrix}$$

where both t_1b_1 and t_2b_2 are larger than 1/2, we compute the limiting covarian matrix to be

$$(5.21) \quad \sum_{m=1}^{n} m^{-2}T^2 \prod_{j=m+1}^{n} (I - j^{-1}TB)^2\pi = \begin{pmatrix} \dfrac{t_1^2}{(2t_1 b_1 - 1)} & 0 \\ 0 & \dfrac{2t_2^2}{(2t_2 b_2 - 1)} \end{pmatrix}$$

Choosing $t_1 = 1/b_1$ and $t_2 = 1/b_2$ will minimize the entries in the matrix in (5.21). Thus, if $b_1 \neq b_2$ we can do better by using $\{n^{-1}T\}$ than by using $\{An^{-1}\}$ since using $\{An^{-1}\}$ would correspond to the case where $t_1 = t_2$.

6. Concluding Remarks. In Sections 3, 4, and 5 we have restricted ourselves to sequences $\{a_n\}$ of the type $a_n = An^{-1}$. It is clear that arguments like the ones presented above can be given for cases where a_n is chosen to be something other than An^{-1} e.g., $a_n = An^{\epsilon-1}$. Due to Examples 1 and 2 however, the results of the previous sections are not likely to be improved very much by using these different sequences. Indeed, for the Robbins-Monro procedure it was shown in [5], Section 7 that under some restrictions, the Robbins-Monro procedure with $a_n = An^{-1}$ for a certain choice of A is optimal in the sense that it is asymptotically minimax for many weight functions. We may remark that this optimum property can be extended with no difficulty to the multi-dimensional Robbins-Monro procedure.

In [4] Burkholder considers somewhat more general processes than considered here in the sense that he permits $M(X_n)$ and $Z(X_n)$ to depend on n as well as X_n. With some modifications of the assumptions we have made this situation can be treated using the methods of Sections 3 and 4. Procedures given by Burkholder for locating points of inflection of a regression function and for finding the maximum of a density function can also be treated using our methods.

It is sometimes of interest to study the asymptotic behavior of $M(X_n) - \alpha$ for the Robbins-Monro procedure and of $M(X_n) - \alpha_0$ for the Kiefer-Wolfowitz procedure. It is easy to see that results about the asymptotic distribution of these quantities can be obtained from the results about the asymptotic distribution of $X_n - \theta$.

REFERENCES

[1] H. D. Block, "*On stochastic approximation*," ONR Report.

[2] J. R. Blum, "Approximation methods which converge with probability one," *Ann Math. Stat.*, Vol. 25 (1954), pp. 382–386

[3] J. R. Blum, "Multidimensional stochastic approximation procedures," *Ann Math. Stat.*, Vol. 25 (1954), pp. 737–744.

[4] D. L. Burkholder, "On a class of stochastic approximation procedures," *Ann Math Stat.*, Vol. 27 (1956), pp. 1044–1059

[5] K. L. Chung, "On a stochastic approximation method," *Ann. Math Stat.*, Vol. 25 (1954), pp 463–483

[6] C. Derman, "An application of Chung's lemma to the Kiefer-Wolfowitz stochastic approximation procedure," *Ann. Math. Stat*, Vol. 27 (1956), pp. 529–532

[7] A. Dvoretzky, "On stochastic approximation," *Proc of the Third Berkeley Symposium*, Vol 1, pp 39–56.

[8] J. L. Hodges, Jr. and E. L. Lehmann, "Two Approximations to the Robbins-Monro Process," *Proc. of the Third Berkeley Symposium*, vol. 1, pp. 95–104.

[9] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp 462–466.

[10] M. Loeve, "*Probability theory*," New York (1955).

[11] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat*, Vol. 22 (1951), pp 400–407.

[12] J. Wolfowitz, "On stochastic approximation methods," *Ann Math Stat*, Vol 27 (1956), pp. 1151–1155

ADMISSIBILITY FOR ESTIMATION WITH QUADRATIC LOSS[1]

By Samuel Karlin

Stanford University

0. Introduction. In dealing with estimation of a single unknown parameter, the criteria most commonly employed in evaluating the worth of given estimates is to make comparisons of the expected square deviation of the estimates from the true value. Suppose on the basis of an observation x (or series of observations) on a distribution $P(x, \omega)$ of the form $\int_{-\infty}^{x} p(\xi, \omega)\, d\mu(\xi)$ depending on an unknown parameter ω it is desired to estimate some function $h(\omega)$. The quantity $p(x, \omega)$ may be regarded as the density of $P(x, \omega)$ with respect to the completely additive measure μ. A non-randomized estimate of $h(\omega)$ is described by a function of the observations $a(x)$, and when the error of an estimate is evaluated in terms of quadratic loss, the expected risk for the estimate $a(x)$ when the true parameter value is ω is calculated by means of the formula

$$\rho(\omega, a) = \int (a(x) - h(\omega))^2 p(x, \omega)\, d\mu(x). \tag{1}$$

The object is to select the estimate a which minimizes (1) in some sense. The fact that the statistician may restrict attention only to non-randomized estimates is due to the convexity property of the loss function ([1], p. 294; [2], p. 4.3). The justification of the quadratic loss as a measure of the discrepancy of an estimate derives from the following two characteristics: (i) in the case where the $a(x)$ represents an unbiased estimate of $h(\omega)$, (1) may be interpreted as the variance of $a(x)$ and, of course, fluctuation as measured by variance is very traditional in the domain of classical estimation; (ii) from a technical and mathematical viewpoint square error lends itself most easily to manipulation and computations.

Principles used to determine a particular estimate which accomplishes appropriate optimizations are related to the minimax criteria, Bayes procedures, unbiased uniformly minimum variance estimates, etc. However, one prerequisite universally acceptable as desirable for statistical procedures is the property of admissibility. An estimate a is said to be admissible if there exists no other estimate a^* such that $\rho(\omega, a^*) \leqq \rho(\omega, a)$ with inequality for some ω. In other words, an estimating procedure is admissible if it cannot be uniformly improved upon in terms of risk by any other procedure. Certainly, no estimate should be used if we can do better by a different estimate—whatever the true state of nature. It would, therefore, be of considerable interest to establish the admissibility of some of the standard estimates employed in practice.

Received May 15, 1957.

[1] This work was sponsored by the Office of Naval Research under Contract Nonr-225(21) (NR-042-993). Reproduction in whole or in part is permitted for any purpose of the United States Government.

A more ambitious undertaking would be to try to characterize all possible admissible estimates for the case of square error. This appears to be an almost insurmountable task. On the other hand, it is relatively easy to determine complete classes of procedures for many parametric problems. In fact, whenever the density $p(x, \omega)$ possesses a monotone likelihood ratio, all possible monotone functions $a(x)$ constitute an essentially complete class of estimating procedures [3] Nevertheless, for any multi-action problem, which includes in particular estimation, it is known that many of the members of a complete class need not be admissible [3], [4]. Furthermore, we have found that admissibility is tied very closely to the order of growth of the loss functions. Square error falls into the category which admits many monotone inadmissible estimates. For absolute error, in contrast, the likelihood that one of the usual estimates is admissible seems to be greater.

Since the general question of resolving admissibility of all estimates measured with respect to quadratic loss function is intrinsically difficult, it seems worth while to concentrate on the investigation of whether some of the most commonly employed classical estimates are admissible.

In this paper we study the problem of admissibility of the usual estimates for three important classes of distributions.

The first class of distributions comprises the exponential family where $p(x, \omega) = \beta(\omega)e^{x\omega}$. The family $a_\gamma(x) = \gamma x$ is considered as possible estimators for $h(\omega) = \frac{-\beta'(\omega)}{\beta(\omega)} = E_\omega(x) = \beta(\omega)\int xe^{x\omega}\, d\mu(x)$ Usually x represents a sufficient statistic based on several observations coming from an exponential distribution. The problem examined in general is whether γx is an admissible estimate of $E_\omega(x)$ measured in terms of quadratic loss. The parameter ω is taken to vary over its natural range Ω consisting of all ω for which $\int e^{x\omega}\, d\mu(x) < \infty$. It is well known that the natural range Ω is an interval which may be finite or infinite. In the case where $\Omega = (-\infty, \infty)$, it has been shown that $a_1(x) = x$ is admissible (see [4] and [5]). The method of proof in both references rests heavily upon the use of the Cramér-Rao inequality and associated differential inequalities. The fact that x is an unbiased estimate of $E_\omega(x)$ seems also to play a fundamental role in this proof. It seems difficult to perceive the meaning behind the analysis and the reasons why things work. In Section 1 we develop a direct proof of this fact. Our methods yield the further interesting and possibly surprising result that γx for any γ satisfying $0 < \gamma \leqq 1$ is an admissible estimate of $E_\omega(x)$ whenever μ possesses positive measure in the regions $x \geqq 0$ and $x \leqq 0$ and $\Omega = (-\infty, \infty)$. On the other hand, for any $\gamma > 1$, γx is not admissible. In view of the fact that any contraction of x (γx, $0 < \gamma \leqq 1$) is admissible it seems surprising that in practice one always uses the extreme estimate of this kind. The criteria of unbiasedness traditionally has dominated the choice of an estimate. Yet we find in several types of estimation problems that this feature of biasing the estimate by scaling it downward is necessary to achieve admissibility. We shall elaborate later on this phenomenon.

If the natural range Ω of ω is not the full infinite interval, then the full determination of the problem of admissibility of γx appears to be complicated. For the special case where

$$p(x, \omega) = \begin{cases} \dfrac{\omega^{\alpha} x^{\alpha-1} e^{-\omega x}}{\Gamma(\alpha)}, & x \geqq 0, \\ 0, & x < 0, \end{cases}$$

for which $\Omega = (0, \infty)$ we find that of all estimates of the form γx there exists a single admissible member in this class, namely $\gamma = [\alpha/(\alpha + 1)]$, which is a biased estimate of $E_{\omega}(x) = \dfrac{\alpha}{\omega}$ (see [4]).

When ω naturally ranges over a finite interval, the problem of admissibility is even more difficult. The analysis seems to depend on the rate at which $\beta(\omega)$ tends to zero as ω approaches its boundary. For example, it is shown later that, if $p(x, \omega) = \beta(\omega)e^{x\omega}(e^{-|x|}/2)$ for which $\Omega = (-1, 1)$ and $\beta(\omega) = 1 - \omega^2$, then all estimates γx $(0 < \gamma \leqq \frac{1}{2})$ are admissible estimates of $E_{\omega}(x)$ while for any other $\gamma > \frac{1}{2}, \gamma x$ may be uniformly improved upon in terms of risk. In general, the possible values of γ for which γx is admissible appears to be very sensitive to the explicit measure $d\mu(x)$ of the exponential family and generally consists of a subinterval of the unit interval.

The following general result concerning admissibility of γx is the assertion of Theorem 1 of Section 1: if $\beta^{-\lambda}(\omega)$ is not integrable in the neighborhood of both boundaries of Ω, then $[1/(\lambda + 1)]x$ is an admissible estimate of $E_{\omega}(x)$. This includes as special cases all previously known results in this direction.

Admissibility is next investigated for the class of distributions where

$$p(x, \omega) = \begin{cases} q(\omega)r(x), & 0 \leqq x \leqq \omega, \\ 0, & x > \omega \text{ or } x < 0, \end{cases} \quad \text{and} \quad d\mu(x) = dx.$$

r is a positive function of x and $q(\omega)$ represents a normalizing constant. This includes, in particular, extremal distributions arising from the uniform density; e.g., $r(x) = nx^{n-1}$, $n \geqq 1$, and $q(\omega) = 1/\omega^n$. We assume in what follows that $r(x)$ is such that the integral $\int_0^{\infty} r(x)\, dx$ diverges. This requires that the normalizing factor $q(\omega)$ approach zero as ω increases to infinity. In dealing with the estimation problem it is convenient to consider estimates of $1/[q^{\alpha}(\omega)]$, $\alpha > 0$, a strictly monotone increasing function of ω. Again we limit attention to estimates which are functions of a single observation x. This in fact is justifiable in every sense whenever the observation x summarizes a sufficient statistic. For example, if $x_1, \cdots, x_n$ represent independent observations from a uniform density spread on the interval $(0, \omega)$, then $\max_{1 \leqq r \leqq n}(x_r) = y$ possesses a density of the form described above, where $r(y) = ny^{n-1}$, and the justification of basing estimates of ω solely on y is manifestly clear.

Although an unbiased estimate of $h(\omega) = 1/2q(\omega)$ is $a(x) = 1/q(x)$, the only admissible estimate of the form $\gamma[1/2q(x)]$, γ a constant, is obtained for the unique value $\gamma = \frac{3}{4}$. Thus, the characteristic phenomenon appears once again

to the effect that admissible estimates are obtained provided the estimate is biased by scaling downward. The same is true when treating the problem of estimating the function $h(\omega) = [1/q(\omega)]^{\alpha}$ with $\alpha > 0$. Analogous results are also valid for the class of distributions

$$p(x, \omega) = \begin{cases} q(\omega)r(x), & x > \omega' \\ 0, & x < \omega, \end{cases}$$

of which

$$p(x, \omega) = \begin{cases} e^{-(x-\omega)}, & x > \omega, \\ 0, & x < \omega, \end{cases}$$

is a typical example.

A possible source of explanation for the excessive uses of the principle of unbiasedness as a basis for selecting one estimate in preference to another may be due to the following considerations: First, a familiar theorem due to Blackwell states that within the collection of all unbiased estimates there exists a uniformly minimum variance unbiased estimate [6]. [This is the case if the family of densities generated by the various parameters is large enough in the sense of forming a "complete family" ([2], p. 3.6.)] This certainly lends importance and some cognizance to the consideration of unbiased estimates. Second, in considering asymptotic or large sample theory, it is found that consistent estimators are, for large sample size, nearly unbiased. For these two reasons, a tradition demanding an estimate be unbiased regardless of the sample size has become acceptable practice. From the point of view of admissibility this is almost universally the wrong estimate to use. We find the desire and need to bias an estimate to insure admissibility.

The third group of distributions studied from the point of view of estimation is related to the important translation parameter problem. The underlying density is assumed known except for a location parameter; that is, $p(x, \omega) = p(x - \omega)$ and we wish to estimate ω. In order for the problem to possess the proper invariance structure we further suppose that $d\mu(x) = dx$, $\int \xi p(\xi)\, d\xi = 0$, and $\int \xi^2 p(\xi)\, d\xi < \infty$. Consequently, we readily observe that for the case of a single observation, x is an unbiased estimate of ω. In the present context the relevance and justification of using the estimate rests primarily on its characteristics of invariance with respect to translations and only incidentally on the property of unbiasedness. With further slight conditions we establish that x is an admissible estimate of ω.

For the situation of several independent observations $x_1, x_2, \cdots, x_n$ the minimum variance invariant estimate is the familiar Pitman estimate

$$(2) \quad a^*(x) = x_1 - \frac{\int \theta p(\theta)p(x_2 - x_1 + \theta)p(x_3 - x_1 + \theta)\cdots p(x_n - x_1 + \theta)\, d\theta}{\int p(\theta)p(x_2 - x_1 + \theta)p(x_3 - x_1 + \theta)\cdots p(x_n - x_1 + \theta)\, d\theta}$$

which represents the multi-observation analogue of the estimate x [5]. If the density $p(\xi)$ is assumed to possess a sufficient number of moments, the expression for $a^*(x)$ is well-defined. Again, subject to sufficient smoothness requirements, we will show that $a^*(x)$ is an admissible estimate of ω. A special case of this result where the parameter and observation both traverse the set of integers was discussed by Blackwell [7]. He demonstrated in this case that $a^*(x)$ is admissible whenever $p(\xi)$ vanishes outside a finite interval. He also showed that without some limitations on the nature of the density p the admissibility of $a^*(x)$ is not generally valid. In connection with the translation parameter problem, a notion of local admissibility is also examined; this notion may possess a greater degree of applicability than indicated in the present context.

The method of analysis in all three cases revolves about an inversion process which we proceed to explain in formal terms. Suppose it is desired to establish that $a(x)$ is an admissible estimate of $h(\omega)$ with respect to the loss function measured by square deviation. Assume the contrary that $b(x)$ is an estimating procedure which improves upon $a(x)$. This states that the inequality

$$\int [b(x) - h(\omega)]^2 p(x, \omega)\, d\mu(x) \leqq \int [a(x) - h(\omega)]^2 p(x, \omega)\, d\mu(x)$$

must be true for all ω. Therefore

$$(3) \qquad \int [b(x) - a(x)]^2 p(x, \omega)\, d\mu(x) \leqq 2 \int [a(x) - b(x)][a(x) - h(\omega)] p(x, \omega)\, d\mu(x)$$

also holds for all ω. In order to demonstrate that $a(x)$ is admissible, it is enough to show that the truth of (3) is only possible provided $b(x) = a(x)$ almost everywhere with respect to μ. Suppose it is possible to construct a monotone increasing function $F(\omega)$, not necessarily bounded, with the property that

$$\int h(\omega) p(x, \omega)\, dF(\omega) = a(x) \int p(x, \omega)\, dF(\omega).$$

Provided that all operations performed are legitimate, it follows that after integrating (3) with respect to dF and interchanging the order of integration

$$\int [b(x) - a(x)]^2 \left[\int p(x, \omega)\, dF(\omega) \right] d\mu(x) \leqq 0.$$

This implies, essentially, the desired result. Throughout what follows, we develop sufficient machinery to justify this formalism. The method may be applied to numerous other kinds of admissibility questions which are not studied in the present paper.

This formalism can also be related to the concept of the optimal Bayes procedure. If $F(\omega)$ represents a bona fide distribution and our objective is to obtain the Bayes estimate of $h(\omega)$ with respect to quadratic loss for $F(\omega)$, then it is a

known fact that the best estimate is given by the expression

$$a(x) = \frac{\int h(\omega)p(x, \omega)\, dF(\omega)}{\int p(x, \omega)\, dF(\omega)}$$

(see [1], p. 299).

Unfortunately, in all cases we are concerned with the relevant $F(\omega)$ turns out to be a non-finite measure. One could then alternatively try to approach $F(\omega)$ by a sequence of distributions such that the corresponding estimates converge to the desired $a(x)$. Such a method of analysis for admissibility was proposed and exploited by Lehmann and Blyth ([2], Section 4 4; [8]). The present results might be viewed as a refinement of this idea.

The extensions of these results and method to the analogous sequential estimation problem will be published subsequently.

Finally, we wish to express our thanks to Mr. Rupert Miller for his help in the writing of this manuscript.

1. Exponential family. In this section the random variable X will be assumed to be distributed according to the probability density $dF_\omega(x) = \beta(\omega)e^{\omega x}\, d\mu(x)$ μ is a σ-finite measure defined on the real line, and ω, the unknown state of nature, belongs to the set $\Omega = \{\omega \mid \int_{-\infty}^{\infty} e^{\omega x}\, d\mu(x) < \infty\}$ which is an interval of the real line. Let $\bar{\omega}$ and $\underline{\omega}$ be the upper and lower endpoints of Ω, respectively. $\bar{\omega}$ and $\underline{\omega}$ may or may not belong to Ω, and in some cases $\bar{\omega} = +\infty$, $\underline{\omega} = -\infty$. The problem for consideration is the estimation of the quantity $\theta(\omega) = E_\omega(x) = -\beta'(\omega)/\beta(\omega)$ from a single observation x on X. There is no loss of generality in restricting our attention to the case of a single observation for, as is well-known, a sufficient statistic for n observations from an exponential distribution is the sum of the observations whose distribution is also a member of the exponential family ([1], p. 221).

Admissible estimates of $\theta(\omega)$ will be derived for the different cases depending on the structure of Ω. We shall consider only classical type estimates of the form $\gamma x = a_\gamma(x)$ where γ is a positive constant. The value $\gamma = 1$ provides the unique unbiased estimate of $E_\omega(x)$ within this family $a_\gamma(x)$.

The only estimate ordinarily considered is $a_1(x) = x$ and this appears to be due to the influence the concept of unbiasedness has had on statistical theory and practice (see our discussion in the introduction). Square error as a measure of the value of an estimate has been tacitly associated also with the principle of unbiasedness. Nevertheless, we shall find that from the point of view of admissibility it is frequently preferred to bias the estimate. Hodges and Lehmann [4] demonstrated the admissibility of $a_1(x) = x$ for a few scattered examples. Girshick and Savage [5] showed that provided $\Omega = (-\infty, \infty)$, x is admissible. Our results cover a substantially larger subclass of the full exponential family for the whole set of estimates $a_\gamma(x)$.

In view of the relations

$$\beta(\omega) \int x^2 e^{\omega x}\, d\mu(x) = \frac{2[\beta'(\omega)]^2 - \beta(\omega)\beta''(\omega)}{\beta^2(\omega)}$$

$$\beta(\omega) \int x e^{\omega x}\, d\mu(x) = -\frac{\beta'(\omega)}{\beta(\omega)},$$

we obtain that

$$\beta(\omega) \int [\gamma x - \theta(\omega)]^2 e^{x\omega}\, d\mu(x) = \gamma^2 \left[\frac{2[\beta'(\omega)]^2 - \beta(\omega)\beta''(\omega)}{\beta^2(\omega)}\right] - 2\gamma \frac{[\beta'(\omega)]^2}{\beta^2(\omega)} + \frac{[\beta'(\omega)]^2}{\beta^2(\omega)}. \tag{4}$$

For each ω in Ω the minimum of the quadratic expression in γ is achieved uniquely for the value

$$\gamma_\omega = \frac{1}{1 + \dfrac{(\beta'(\omega))^2 - \beta(\omega)\beta''(\omega)}{(\beta'(\omega))^2}} = \frac{\left(\int x e^{x\omega}\, d\mu\right)^2}{\left(\int e^{x\omega}\, d\mu\right)\left(\int x^2 e^{x\omega}\, d\mu\right)}. \tag{5}$$

But $(\beta'(\omega))^2 - \beta(\omega)\beta''(\omega) = \beta^2(\omega)\sigma_x^2 > 0$ (σ_x^2 = variance of x) so that $0 < \gamma_\omega \leqq 1$. This inequality satisfied by γ_ω can also be deduced as a consequence of the Schwarz inequality on inspection of the second formula for γ_ω. It follows for any $\gamma > 1$, $\rho(\omega, a_\gamma) = \beta(\omega)\int [\gamma x - \theta(\omega)]^2 e^{x\omega}\, d\mu(x)$ is strictly increasing in γ for all ω. Consequently, if γ' is chosen satisfying $1 \leqq \gamma' < \gamma$, then $\rho(\omega, a_{\gamma'}(x)) < \rho(\omega, a_\gamma(x))$ for all ω in Ω and therefore $a_\gamma(x)$ is not admissible. This argument can be extended as follows: Suppose $[(\beta'(\omega))^2 - \beta(\omega)\beta''(\omega)]/(\beta'(\omega))^2$ ranges between L and L' ($L < L'$) as ω traverses the interval $(\underline{\omega}, \bar{\omega})$. Then γ_ω lies in the range $(1/(1 + L'), 1/(1 + L)) = I$ and for any $\gamma > 1/(1 + L)$ the same reasoning shows that $a_\gamma(x)$ is not admissible. Whenever Ω is not the full infinite interval for many circumstances $1/(1 + L) < 1$ and x is therefore not admissible. The converse implication is not valid. That is, if γ lies interior to I, then it is not necessarily true that $a_\gamma(x)$ is admissible. A counter-example may be provided as follows: Suppose the measure μ is such that it spreads its entire mass throughout the interval $1 \leqq x \leqq 2$. Then, $\theta(\omega) = E_\omega(x)$ likewise traverses the interval [1, 2] as ω varies over the set $\Omega = (-\infty, \infty)$. No estimate of the form γx $(0 < \gamma < 1)$ can be admissible since this entails estimating $\theta(\omega)$ as less than one with positive probability. Whenever the observed $x < (1/\gamma)$, which occurs with positive probability, an immediate improvement of the proposed estimate γx is obtained by estimating $\theta(\omega)$ as 1 in that range. This emphasizes the fact that an estimate $a_{\gamma_0}(x)$, admissible with respect to all estimates $a_\gamma(x)$, need not be universally admissible.

We direct attention to the question of admissibility for $a_\gamma(x)$ where γ is in I. Suppose $g(x)$ is an estimate which satisfies $\rho(\omega, g) \leqq \rho(\omega, a_\gamma)$ for all ω. This in-

equality may be reduced to the form

$$\int_{-\infty}^{\infty} [g(x) - \gamma x]^2 \beta(\omega) e^{x\omega}\, d\mu(x) \leqq 2 \int_{-\infty}^{\infty} [\gamma x - g(x)][\gamma x \beta(\omega) + \beta'(\omega)] e^{x\omega}\, d\mu(x).$$

Let $dF(\omega) = \beta^{\lambda}(\omega)\, d\omega$ for constant $\lambda \neq -1$, and let $a, b \in \Omega$, $a < b$. Also define $T(\omega) = \int_{-\infty}^{\infty} [g(x) - \gamma x]^2 \beta(\omega) e^{x\omega}\, d\mu(x)$. Then,

$$(6)\quad \begin{aligned} &\int_a^b \beta^{\lambda}(\omega) T(\omega)\, d\omega \\ &\leqq 2 \int_a^b \beta^{\lambda}(\omega) \left\{ \int_{-\infty}^{\infty} [\gamma x - g(x)][\gamma x \beta(\omega) + \beta'(\omega)] e^{x\omega}\, d\mu(x) \right\} d\omega \\ &= 2 \int_{-\infty}^{\infty} [\gamma x - g(x)] \left[\frac{\beta^{\lambda+1}(b) e^{xb}}{\lambda + 1} - \frac{\beta^{\lambda+1}(a) e^{xa}}{\lambda + 1} \right] d\mu(x) \\ &+ 2 \int_{-\infty}^{\infty} [\gamma x - g(x)] \left[\gamma x - \frac{1}{\lambda + 1} x \right] \left[\int_a^b \beta^{\lambda+1}(\omega) e^{x\omega}\, d\omega \right] d\mu(x). \end{aligned}$$

Suppose $\gamma = 1/(\lambda + 1)$. Then, the last term in (6) vanishes, and by a proper application of Schwarz's inequality, (6) becomes (for $\gamma = 1/(\lambda + 1)$)

$$(7)\quad \begin{aligned} &\int_a^b \beta^{\lambda}(\omega) T(\omega)\, d\omega \\ &\leqq \frac{2}{\lambda + 1} \sqrt{\beta^{\lambda}(b)}\, \sqrt{T(b)\beta^{\lambda}(b)} + \frac{2}{\lambda + 1} \sqrt{\beta^{\lambda}(a)}\, \sqrt{T(a)\beta^{\lambda}(a)}. \end{aligned}$$

Let c be an interior point of Ω. Suppose $\int_c^b \beta^{-\lambda}(\omega)\, d\omega \to +\infty$ as $b \to \bar{\omega}$ and $\int_a^c \beta^{-\lambda}(\omega)\, d\omega \to +\infty$ as $a \to \underline{\omega}$. Then it follows that (see Cases 1 and 2 below) $T(\omega) = 0$, a.e. But this requires that $g(x) = [1/(\lambda + 1)]x$, a.e.; that is, the estimate $x/(\lambda + 1)$ is an admissible estimate.

Case 1.

$$\underline{\lim}_{b \to \bar{\omega}} \beta^{\lambda}(b) \sqrt{T(b)} = \Delta > 0.$$

Fix a and let $H(b) = \int_a^b \beta^{\lambda}(\omega) T(\omega)\, d\omega$. By virtue of (7) we can find an appropriate constant $C > 0$ such that for b sufficiently close to $\bar{\omega}$,

$$H(b) \leqq C \sqrt{\beta^{\lambda}(b)}\, \sqrt{H'(b)}.$$

This yields by transposition and integration

$$C^2 \left[\frac{1}{H(b_1)} - \frac{1}{H(b_2)} \right] \geqq \int_{b_1}^{b_2} \beta^{-\lambda}(b)\, db,$$

where b_1, b_2 are chosen so that $b_1 < b_2$, $H(b_1) > 0$. As $b_2 \to \bar{\omega}$ the right-hand side tends to $+\infty$ and the left-hand side remains bounded—which is impossible. Thus, Case 1 cannot occur.

CASE 2.

$$\lim_{b \to \bar{\omega}} \beta^{\lambda}(b) \sqrt{T(b)} = 0.$$

Let $G(a) = \int_a^{\bar{\omega}} \beta^{\lambda}(\omega) T(\omega)\, d\omega$. By (7) and the assumption of Case 2 it follows that $G(a) \leqq [2/|\lambda + 1|] \sqrt{\beta^{\lambda}(a)} \sqrt{-G'(a)}$. Suppose there exists an a_0 such that $G(a_0) > 0$. Then

$$\left(\frac{2}{\lambda + 1}\right)^2 \left[\frac{1}{G(a_0)} - \frac{1}{G(a_1)}\right] \geqq \int_{a_1}^{a_0} \beta^{-\lambda}(a)\, da, \tag{8}$$

where $a_1 < a_0$. As $a_1 \to \underline{\omega}$ the right-hand side tends to $+\infty$ while the left-hand side remains bounded. This is impossible so $G(a) \equiv 0$, which implies $T(\omega) = 0$, a.e. We summarize the conclusions in the statement of a theorem.

THEOREM 1. *Let $p(x, \omega) = \beta(\omega) e^{x\omega}$ describe the density of the exponential family with respect to a measure μ. If*

$$\int_c^b \beta^{-\lambda}(\omega)\, d\omega \to +\infty \qquad \text{as } b \to \bar{\omega} \tag{i}$$

and

$$\int_a^c \beta^{-\lambda}(\omega)\, d\omega \to +\infty \qquad \text{as } a \to \underline{\omega}, \tag{ii}$$

where c is an interior point of $\Omega = (\underline{\omega}, \bar{\omega})$, then $[1/(\lambda + 1)]x$ is an admissible estimate of $\theta(\omega) = E_{\omega}(x)$.

This theorem subsumes as special consequences all previous known results in this direction (see [4] and [5]). We record several specific applications of this theorem of special interest.

I. If $\Omega = (-\infty, \infty)$ and μ possesses positive measure in each of the intervals $(0, \infty)$ and $(-\infty, 0)$, then $a_{\gamma}(x) = \gamma x$ for each $0 < \gamma \leqq 1$ is admissible. In fact, the assumptions imply that

$$\beta(\omega) = \frac{1}{\int e^{x\omega}\, d\mu(x)}$$

converges to zero as $|\omega| \to \infty$. Consequently (i) and (ii) hold for each $\lambda \geqq 0$.

II. If $\Omega = (-\infty, \infty)$ and there exists positive probability of observing the value zero, then $a_{\gamma}(x) = \gamma x$ for each $0 < \gamma \leqq 1$ is admissible. The proof follows readily from Theorem 1 since $\beta(\omega)$ is bounded above.

III. If $\Omega = (-\infty, \infty)$ with no further conditions specified as to the nature of μ, then at least $a_1(x) = x$ is an admissible estimate of $\theta(\omega)$. This is so since the hypotheses of Theorem 1 are satisfied for $\lambda = 0$.

IV. If $p(x, \omega) = \dfrac{(-\omega)^{\alpha}}{\Gamma(\alpha)} x^{\alpha - 1} e^{\omega x}$ for x positive where $\alpha > 0$ is fixed and ω ranges over $\Omega = (-\infty, 0)$, then $\beta(\omega) = (-\omega)^{\alpha}$ and $\theta(\omega) = -\alpha/\omega$. The unique

value of λ satisfying (i) and (ii) is equal to $1/\alpha$. Consequently, $\alpha x/(\alpha + 1)$ is the only admissible estimate of $-(\alpha/\omega)$ of the form γx. In the case of n observations $x_1, x_2, \cdots, x_n$ with x_i independently normally distributed, mean 0 and variance σ^2, this result reduces to the well-known fact that

$$\theta(x) = [1/(n+2)] \sum_1^n x_i^2$$

is an admissible estimate of σ^2. The interval I in this case also reduces to a unique point.

V. If $d\mu(x) = \frac{1}{2}e^{-|x|}$, then $\beta(\omega) = 1 - \omega^2$ and the hypotheses of Theorem 1 are satisfied with $\lambda \geqq 1$. It follows that $a_\gamma(x) = \gamma x$ is admissible for $\gamma \leqq \frac{1}{2}$. Also, in this case $I = (0, \frac{1}{2})$ so that no estimate of the form γx may be admissible for $\gamma > \frac{1}{2}$.

Further examples of similar type involving definite biasing can be cited. In numerous examples calculated where Ω has at least one finite boundary we found that $a_1(x) = x$ is not admissible. We propose a stronger assertion which includes this observation. We state in conjecture that the hypotheses of Theorem 1 are also necessary conditions for the admissibility of the corresponding estimate. This would imply in particular that whenever $\beta(\omega)$ approaches infinity exponentially as ω tends to one of its boundaries no estimate of the form γx can be admissible.

2. Extreme value densities. In this section we consider densities of the form

$$p(x, \omega) = \begin{cases} q(\omega)r(x), & 0 \leqq x \leqq \omega, \\ 0, & \text{otherwise,} \end{cases}$$

where $r(x)$ is assumed to be a positive Lebesgue measurable function of x and $q^{-1}(\omega) = \int_0^\omega r(x)\,dx < \infty$ for ω in $\Omega = (0, \infty)$ We further assume that the monotone decreasing function $q(\omega)$ approaches zero as $\omega \to \infty$, or equivalently $\int_0^\infty r(x)\,dx = \infty$.

The problem examined concerns estimating functions of the form $[1/q(\omega)]^\alpha$, $\alpha > 0$. In determining proper estimators attention is directed only to estimates also of the form $\gamma[1/q(x)]^\alpha = a_\gamma(x)$ where γ is a positive constant. It is reasonable and justifiable to consider only a single observation because of the fact that x ordinarily represents a sufficient statistic.

Since $r(x) = -q'(x)/q^2(x)$ almost everywhere, we find

$$\begin{aligned} \rho(\omega, a_\gamma) &= q(\omega) \int_0^\omega \left[\frac{\gamma}{q^\alpha(x)} - \frac{1}{q^\alpha(\omega)}\right]^2 r(x)\,dx \\ &= \left[\frac{\gamma^2}{2\alpha+1} - \frac{2\gamma}{\alpha+1} + 1\right] \frac{1}{q^{2\alpha}(\omega)}. \end{aligned} \tag{9}$$

Hence, the minimum of the quadratic expression is achieved uniformly with respect to ω for the single choice $\gamma = (2\alpha + 1)/(\alpha + 1)$. For comparison purposes we note that within the family of estimates considered the unbiased estimate of $1/q^\alpha(\omega)$ is $(\alpha + 1)/q^\alpha(x)$. The unbiased estimate can therefore be uni-

formly improved upon in terms of expected risk by applying the bias factor $(2\alpha + 1)/(\alpha + 1)^2 < 1$. We proceed to demonstrate the admissibility of the estimator $[(2\alpha + 1)/(\alpha + 1)1]/q^\alpha(x)$ as an estimate of $1/q^\alpha(\omega)$.

The method of proof follows the same general ideas as used in the preceding section. Suppose $g(x)$ is an estimate which satisfies the property that for all ω

$$\rho(\omega, g) \leqq \rho(\omega, a_\gamma), \qquad \gamma = \frac{2\alpha + 1}{\alpha + 1}.$$

Consequently,

$$(10)\qquad \begin{aligned} a(\omega) &= \int_0^\omega \left(g(x) - \frac{\gamma}{q^\alpha(x)}\right)^2 q(\omega)r(x)\,dx \\ &\leqq 2\int_0^\omega \left[\frac{\gamma}{q^\alpha(x)} - g(x)\right]\left[\frac{\gamma}{q^\alpha(x)} - \frac{1}{q^\alpha(\omega)}\right] q(\omega)r(x)\,dx. \end{aligned}$$

In order to check admissibility for $a_\gamma(x)$ it is enough to show that the only g satisfying this system of inequalities is $g(x) = a_\gamma(x)$, a.e. In view of the formalism indicated in the introduction the aim is to integrate the formula of (10) with respect to an appropriate monotone increasing function in order to cause the right-hand side to vanish. This essentially implies admissibility. Accordingly, we select $dF(\omega) = |q'(\omega)|\,q^\beta(\omega)\,d\omega$ where $\beta = 2\alpha - 1$. Then,

$$(\beta + 2)/(\beta + 2 - \alpha) = \gamma = (2\alpha + 1)/(\alpha + 1).$$

By direct calculation we obtain

$$(11)\qquad \begin{aligned} &\int_\epsilon^\tau a(\omega)\,|q'(\omega)|\,q^\beta(\omega)\,d\omega \leqq \\ &2\int_\epsilon^\tau \left\{\int_0^\omega \left[\frac{\gamma}{q^\alpha(x)} - g(x)\right]\left[\frac{\gamma}{q^\alpha(x)} - \frac{1}{q^\alpha(\omega)}\right] |q'(\omega)|\,q^{\beta+1}(\omega)r(x)\,dx\right\} d\omega \\ &= \frac{2}{\beta + 2 - \alpha}\left\{\int_0^\tau \left[\frac{\gamma}{q^\alpha(x)} - g(x)\right] r(x)q^{\beta+2-\alpha}(\tau)\left[1 - \frac{q^\alpha(\tau)}{q^\alpha(x)}\right] dx\right\} \\ &\quad - \frac{2}{\beta + 2 - \alpha}\left\{\int_0^\epsilon \left[\frac{\gamma}{q^\alpha(x)} - g(x)\right] r(x)q^{(\beta+2-\alpha)}(\epsilon)\left[1 - \frac{q^\alpha(\epsilon)}{q^\alpha(x)}\right] dx\right\}. \end{aligned}$$

Since $q(x) \geqq q(\tau)$ for $x \leqq \tau$, we deduce with the aid of Schwarz's inequality that

$$(12)\qquad \begin{aligned} &q^\alpha(\tau)\int_0^\tau \left|\frac{\gamma}{q^\alpha(x)} - g(x)\right| \sqrt{r(x)q(\tau)}\,\sqrt{r(x)q(\tau)}\left|1 - \frac{q^\alpha(\tau)}{q^\alpha(x)}\right| dx \\ &\qquad \leqq \sqrt{a(\tau)}\,q^\alpha(\tau) = \sqrt{a(\tau)q^\beta(\tau)}\,\sqrt{q(\tau)}. \end{aligned}$$

In a similar way the second integral of (11) has a bound equal to

$$\sqrt{a(\epsilon)q^\beta(\epsilon)}\,\sqrt{q(\epsilon)}.$$

By combining the relations of (11) and (12) and the last stated bound, we

obtain

$$(13)\quad \int_\epsilon^\tau a(\omega)\,|\,q'(\omega)\,|\,q^\beta(\omega)\,d\omega \leqq \frac{2}{\alpha+1} \times\left[\sqrt{a(\tau)q^\beta(\tau)\,|\,q'(\tau)\,|}\,\sqrt{\frac{q(\tau)}{|\,q'(\tau)\,|}} + \sqrt{a(\epsilon)q^\beta(\epsilon)\,|q'(\epsilon)\,|}\,\sqrt{\frac{q(\epsilon)}{|\,q'(\epsilon)\,|}}\right]$$

The analysis proceeds by examining two possible cases

CASE 1.

$$\lim_{\tau\to\infty} \sqrt{a(\tau)\,|\,q'(\tau)\,|\,q^\beta(\tau)}\,\sqrt{q(\tau)/|\,q'(\tau)\,|} = \Delta > 0$$

Fix ϵ and set $H(\tau) = \int_\epsilon^\tau a(\omega)|\,q'(\omega)|\,q^\beta(\omega)\,d\omega$ There exists a constant C such that for sufficiently large τ

$$(14)\qquad H(\tau) \leqq C\sqrt{H'(\tau)}\,\sqrt{\frac{q(\tau)}{q'(\tau)}}.$$

We now show that this relation leads to an absurdity Indeed, squaring the expression of (14) and solving the differential inequality, we deduce that

$$(15)\qquad C^2\left[\frac{1}{H(\beta)} - \frac{1}{H(\alpha)}\right] \leqq \log\frac{q(\beta)}{q(\alpha)},$$

where $\beta > \alpha$ and α sufficiently large As $\beta \to \infty$ the left-hand side of (15) remains bounded while the right-hand side tends to $-\infty$ which is impossible. Thus, Case 1 cannot occur.

CASE 2.

$$\lim_{\tau\to\infty} \sqrt{a(\tau)\,|\,q'(\tau)\,|\,q^\beta(\tau)}\,\sqrt{q(\tau)/|\,q'(\tau)\,|} = 0$$

Let τ tend to $+\infty$ along a sequence $\{\tau_n\}$ for which

$$\lim_{n\to\infty} \sqrt{a(\tau_n)\,|\,q'(\tau_n)\,|\,q^\beta(\tau_n)}\,\sqrt{q(\tau_n)/|\,q'(\tau_n)\,|} = 0$$

Then, by (13)

$$(16)\quad G(\epsilon) = \int_\epsilon^\infty a(\omega)\,|\,q'(\omega)\,|q^\beta(\omega)\,d\omega \leqq \frac{2}{\alpha+1}\sqrt{a(\epsilon)\,|\,q'(\epsilon)\,|\,q^\beta(\epsilon)}\,\sqrt{q(\epsilon)/|q'(\epsilon)|}.$$

Suppose $G(\epsilon_0) > 0$. Then $G(\epsilon) \geqq G(\epsilon_0) > 0$ for $\epsilon \leqq \epsilon_0$. (16) can be written as $G(\epsilon) \leqq [2/(\alpha+1)]\sqrt{-G'(\epsilon)}\sqrt{q(\epsilon)/|\,q'(\epsilon)\,|}$. Transposition of terms in this expression and integration over (ϵ_1, ϵ_0) yields

$$(17)\qquad \left(\frac{2}{\alpha+1}\right)^2\left[\frac{1}{G(\epsilon_1)} - \frac{1}{G(\epsilon_0)}\right] \leqq \log\frac{q(\epsilon_0)}{q(\epsilon_1)}.$$

As $\epsilon_1 \to 0$ the left-hand side remains bounded but the right-hand side tends to $-\infty$ which is an absurdity. Thus the supposition that $G(\epsilon_0) > 0$ for some $\epsilon_0 > 0$

is erroneous, and therefore $G(\epsilon) \equiv 0$. Consequently, $a(\omega) = 0$, a.e., which implies $g(x) = \gamma/q^{\alpha}(x)$, a.e.

We have thus established the truth of

THEOREM 2. *There exists a single admissible estimate of* $1/q^{\alpha}(\omega)$ *of the form* $\gamma/q^{\alpha}(x)$, *and this is given by* $\gamma = (2\alpha + 1)/(\alpha + 1)$.

The following specific application might be of some interest. Let $r(x) = nx^{n-1}$. Then $[(2 + n)/(1 + n)]x$ is an admissible estimate of ω. Furthermore, this is the only admissible estimate which is a multiple of x.

This states that if $x_1, x_2, \cdots, x_n$ represents n independent observations from a rectangular density spread on the interval $(0, \omega)$, then

$$[(n + 2)/(n + 1)] \max_i x_i$$

is an admissible estimate of ω with respect to squared error.

To pinpoint the reasons for the validity of the preceding methodology it seems worth emphasizing that although for any γ it is possible to construct a measure $q^{\beta}(\omega)|\, q'(\omega)|$ which formally gives

$$\frac{\gamma}{q^{\alpha}(x)} = \frac{\int_x^{\infty} q^{\beta+1-\alpha}(\omega)q'(\omega)\, d\omega}{\int_x^{\infty} q^{\beta+1}(\omega)q'(\omega)\, d\omega},$$

nevertheless, the reader will find that it is only possible to justify the formalism for the special choices $\beta = 2\alpha - 1$ and $\gamma = (2\alpha + 1)/(\alpha + 1)$ as we have done. The estimate $[(2\alpha + 1)/(\alpha + 1)]/q^{\alpha}(x)$ of $1/q^{\alpha}(\omega)$, although uniquely admissible with respect to square error, is still not altogether acceptable. It is disturbing to note that the estimate $[(2\alpha + 1)/(\alpha + 1)]/q^{\alpha}(x)$ is very closely tied to the measure of error described by quadratic loss. If the risk function is given by

$$\begin{aligned} \rho(\omega, a_{\gamma}(x)) &= q(\omega) \int_0^{\omega} \left[\frac{\gamma}{q^{\alpha}(x)} - \frac{1}{q^{\alpha}(\omega)}\right]^{2k} r(x)\, dx \\ &= \left\{\sum_{r=0}^{2k} (-1)^r \frac{\gamma^r}{r\alpha + 1} \binom{2k}{r}\right\} \frac{1}{q^{2k\alpha}(\omega)}, \end{aligned}$$

it can be shown that the minimum is achieved uniformly in ω at a value γ_k which strictly varies with k. This implies that an admissible estimate with respect to square error need not be admissible when considered for the error function involving 4th powers. It is found that when $\alpha = 1$ in the case of square error the best estimate of the type $\gamma/q(x)$ is $\frac{3}{2}\, 1/q(x)$ while for the loss function of fourth powers the best estimate is $\gamma^*/q(x)$ where $\gamma^* > 0$ satisfies

$$\tfrac{4}{5}\gamma^3 - 3\gamma^2 + 4\gamma - 2 = 0,$$

which is slightly larger than $\frac{3}{2}$.

We close this section with a brief discussion of the problem of admissibility

for the density

$$p(x, \omega) = \begin{cases} q(\omega)r(x), & x \geqq \omega, \\ 0, & \omega_0 \leqq x < \omega, \end{cases}$$

where $r(x)$ is a positive measurable function of x and $q^{-1}(\omega) = \int_\omega^\infty r(x)\,dx < \infty$ for ω in $\Omega = (\omega_0, \infty)$.

One important such example is furnished by taking $r(x) = e^{-x}$, $q(\omega) = e^{\omega}$, and $\Omega = (-\infty, \infty)$. Another example is obtained by setting $r(x) = 1/x^{\delta}$, $\delta > 1$ and $\omega_0 = 0$. As before our problem is to estimate the quantity $1/q^{\alpha}(\omega)$ by using estimates of the form $\gamma/q^{\alpha}(x)$. We assume in what follows that $q(\omega_0) = 0$ or equivalently $\int_{\omega_0}^\infty r(x)\,dx = \infty$.

THEOREM 3. *If*

$$p(x, \omega) = \begin{cases} q(\omega)r(x), & x \geqq \omega, \\ 0, & \omega_0 \leqq x < \omega, \end{cases}$$

where $q^{-1}(\omega) = \int_\omega^\infty r(x)\,dx$ *and* $q(\omega_0) = 0$, *then* $[(2\alpha + 1)/(\alpha + 1)]/q^{\alpha}(x)$ *is an admissible estimate of* $1/q^{\alpha}(\omega)$ *with respect to quadratic loss*

The proof of Theorem 3 parallels that of Theorem 2 subject to simple obvious modifications and will therefore be omitted

3. Translation parameter problem: single observation. The random variable X is distributed according to the probability density $p(x, \omega) = p(x + \omega)$ where $\omega \in \Omega$ is the unknown state of nature and $p(\xi)$ is a known, fixed density function which satisfies $\int_{-\infty}^\infty \xi p(\xi)\,d\xi = 0$. The analogous problem where X is an integer-valued random variable and the parameter likewise ranges over the set of discrete integers will be discussed later. The problem is to estimate the parameter $-\omega$. If x is the single observed value, then the usual (unbiased, invariant) estimate of $-\omega$ is $\delta(x) = x$. The property of unbiasedness is easily verified and for its relationship to invariance the reader is referred to [1]. The principal goal of this section is to establish the admissibility of this estimate, $\delta(x) = x$, subject to appropriate smoothness conditions

This formulation of the translation parameter problem differs notationally from the customary version. If ω is substituted for $-\omega$, then the familiar form of the problem will emerge. The difference in the formulation of the problem is not significant in any way and on the other hand is helpful in that it leads to a more convenient form for applying theorems on Fourier transforms.

To establish admissibility it is sufficient to show that the inequality

$$\rho(\omega, g) \leqq \rho(\omega, \delta),$$

or equivalently

$$(18) \qquad \int_{-\infty}^\infty [x - g(x)]^2 p(x + \omega)\,dx \leqq 2\int_{-\infty}^\infty [x - g(x)][x + \omega]p(x + \omega)\,dx,$$

implies $g(x) = x$, a.e.

To accomplish this it is necessary to impose the following assumption.

ASSUMPTION I.

$$\int_{-\infty}^{\infty} \xi^2 p(\xi)\, d\xi < \infty, \qquad \int_{-\infty}^{\infty} \xi^2 p^2(\xi)\, d\xi < \infty, \quad \text{and} \quad \int_{-\infty}^{\infty} \xi p(\xi)\, d\xi = 0.$$

The meaning and relevance of the last condition was discussed above. The first integrability requirement is indispensible in order that (18) define a meaningful relationship. The second finiteness condition represents a slight further restriction beyond that of the first integral. For instance, the second integrability condition would be an immediate consequence of the first integrability condition and boundedness of the density $p(\xi)$.

We further assume initially that we deal only with alternative estimates $g(x)$ satisfying $| g(x) - x | \leqq M < \infty$. The nature of this restriction is considerably milder than might appear at first glance. It will later be shown that this constraint may be completely eliminated or, equivalently, we will show the only estimates for which (18) is possible must satisfy this restraint.

Unless stated to the contrary we suppose hereafter that Assumption I and the boundedness requirement on competing estimates are satisfied.

LEMMA 1. *If* $\rho(\omega, g) \leqq \rho(\omega, \delta)$ *for all* ω, *then* $\int_{-\infty}^{\infty} (g(x) - x)^2\, dx < \infty$.

PROOF. Define $\Phi(u) = \int_{-\infty}^{u} \xi p(\xi)\, d\xi$. Then

$$\int_{-n}^{n} \left[\int_{-\infty}^{\infty} [x - g(x)]^2 p(x + \omega)\, dx \right] d\omega$$

$$\leqq 2 \int_{-\infty}^{\infty} | x - g(x) |\, dx \left| \int_{-n}^{n} (x + \omega) p(x + \omega)\, d\omega \right| \tag{19}$$

$$\leqq 2M \int_{-\infty}^{\infty} | \Phi(x + n) - \Phi(x - n) |\, dx \leqq 4M \int_{-\infty}^{\infty} (-\Phi(u))\, du$$

as $-\Phi(u)$ is positive. But,

$$\left| u \int_{-\infty}^{u} \xi p(\xi)\, d\xi \right| \leqq \int_{-\infty}^{u} \xi^2 p(\xi)\, d\xi \qquad \text{as } u \to -\infty$$

and

$$\left| u \int_{-\infty}^{u} \xi p(\xi)\, d\xi \right| = \left| u \int_{u}^{\infty} \xi p(\xi)\, d\xi \right| \leqq \int_{u}^{\infty} \xi^2 p(\xi) d\xi \quad \text{as } u \to +\infty$$

so

$$| u\Phi(u)| \to 0 \qquad \text{as} \qquad | u | \to \infty.$$

Hence, integration by parts yields

$$\int_{-\infty}^{\infty} (-\Phi(u))\, du = \int_{-\infty}^{\infty} \xi^2 p(\xi)\, d\xi < \infty.$$

Allowing n to tend to infinity in (19) after interchanging the order of integration produces the desired result.

THEOREM 4. *If $\rho(\omega, g) \leqq \rho(\omega, \delta)$ for all ω, then $g(x) = \delta(x)$, a.e.*

PROOF. By Lemma 1 $\theta_1(\xi) = \xi - g(\xi) \in L^2$ so by Plancherel's theorem its Fourier transform $\varphi_1(u)$ is defined and belongs to L^2.

$$\varphi_1(u) \sim \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iu\xi}\theta_1(\xi)\, d\xi.$$

According to Assumption I $\theta_2(\xi) = \xi p(\xi) \in L^2$ so also its Fourier transform $\varphi_2(u)$ exists and

$$\varphi_2(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iu\xi}\theta_2(\xi)\, d\xi$$

(since the integral exists). The function $\theta(\omega) = \int_{-\infty}^{\infty} \theta_1(x)\theta_2(x + \omega)\, dx$, which is essentially a convolution of θ_1 and θ_2, also belongs to L^2. It can be readily verified that its Fourier transform is $\varphi(u) = \varphi_1(-u)\varphi_2(u)(\varepsilon L^2)$.

Since φ_1 and φ_2 both belong to L^2, by Schwarz's inequality

$$\varphi = \varphi_1(-u)\varphi_2(u) \in L^1.$$

By the inversion theorem on Fourier transforms

$$\int_{-\infty}^{\infty} (x - g(x))(x + \omega)p(x + \omega)\, dx = \int_{-\infty}^{\infty} e^{-i\omega u}\varphi_1(-u)\varphi_2(u)\, du \tag{20}$$

for real ω as both sides represent continuous functions, the first by virtue of the fact that $\xi p(\xi)$ is in L^1 and the second since $\varphi_1 \cdot \varphi_2$ is in L^1 as established. If both sides of (20) are integrated from $-n$ to n and the order of integration is reversed on the right-hand side (which is permissible for reasons indicated below)

$$\int_{-n}^{n} \left\{ \int_{-\infty}^{\infty} (x - g(x))(x + \omega)p(x + \omega)\, dx \right\} d\omega = \int_{-\infty}^{\infty} \frac{\varphi_1(-u)\,\varphi_2(u)}{iu} [e^{-inu} - e^{inu}]\, du. \tag{21}$$

To justify the interchange on the right-hand side we observe first that $\varphi_2(0) = 0$ while $\varphi_2'(u) = [i/\sqrt{2\pi}]\int_{-\infty}^{\infty} e^{iu\xi}\xi^2 p(\xi)\, d\xi$ is bounded independently of u. By the mean value theorem $\varphi_2(u)/u = \varphi_2'(\bar{u})$ where $0 \leqq \bar{u} \leqq u$. Thus,

$$\lim_{u \to 0} \varphi_2(u)/u < \infty$$

and $\int_{-\epsilon}^{\epsilon} \varphi_2(u)/u\, du < \infty$ for $\epsilon > 0$. This implies $[\varphi_1(-u)\varphi_2(u)]/u \in L^1$. But the Riemann-Lebesgue theorem asserts that for any $q \in L^1, \int_{-\infty}^{\infty} e^{iu\omega}q(u)\, du \to 0$ as $\omega \to \pm\infty$. Therefore,

$$\lim_{n \to \infty} \int_{-n}^{n} \left\{ \int_{-\infty}^{\infty} (x - g(x))(x + \omega)p(x + \omega)\, dx \right\} d\omega = 0,$$

and on account of (18) we may infer that

$$\int_{-\infty}^{\infty} (g(x) - x)^2\, dx = \int_{-\infty}^{\infty} (g(x) - x)^2 \left\{ \int_{-\infty}^{\infty} p(x + \omega)\, d\omega \right\} dx \leqq 0,$$

which implies $g(x) = x$, a.e.

As mentioned previously, there are two general cases in which the boundedness assumption is satisfied by all g which need be considered.

CASE 1: $$p(\xi)\begin{cases} = 0, & \xi < a, \xi > b, -\infty < a < b < \infty, \\ \geqq 0, & \text{otherwise.} \end{cases}$$

This type of density is fairly general and will occur, for instance, when any distribution is truncated at finite endpoints.

Suppose x is the observed value. Then, because of the form of $p(\xi)$

$$x - b \leqq -\omega \leqq x - a.$$

Any estimate which assumes values outside the interval $[x - b, x - a]$ can be improved upon by an estimate $h(x)$ which satisfies $x - b \leqq h(x) \leqq x - a$ for all x. Intuitively this is clear; a rigorous proof may be readily supplied by the reader. Thus if $\rho(\omega, g) \leqq \rho(\omega, \delta)$ for all ω and g does not satisfy the boundedness assumption, there exists another estimate $h(x)$ such that

$$| h(x) - x | \leqq | a | + | b |$$

and $\rho(\omega, h) \leqq \rho(\omega, g) \leqq \rho(\omega, \delta)$ for all ω. Since this implies $h(x) = x$, a.e., and hence $\rho(\omega, h) \equiv \rho(\omega, \delta)$, $\rho(\omega, g) \equiv \rho(\omega, \delta)$ which implies that δ is admissible.

In addition, note that Assumption I is automatically satisfied in this case.

CASE 2: $p(x, \omega) = p(x + \omega)$ has a monotone likelihood ratio.

LEMMA 2: *If $\rho(\omega, g) \leqq \rho(\omega, \delta)$ for all ω, then there exists a constant C such that*

$$\int_{-\infty}^{\infty} (g(x) - x)^2 p(x + \omega)\, dx \leqq C$$

for all ω (under Assumption I).

PROOF. By Schwarz's inequality

$$\frac{1}{2}\int_{-\infty}^{\infty} (g(x) - x)^2 p(x + \omega)\, dx < \int_{-\infty}^{\infty} (x - g(x))(x + \omega)p(x + \omega)\, dx$$
$$\leqq \left[\int_{-\infty}^{\infty} (g(x) - x)^2 p(x + \omega)\, dx\right]^{\frac{1}{2}} \left[\int_{-\infty}^{\infty} (x + \omega)^2 p(x + \omega)\, dx\right]^{\frac{1}{2}}.$$

It follows easily that

$$\int_{-\infty}^{\infty} (g(x) - x)^2 p(x + \omega)\, dx \leqq 4 \int_{-\infty}^{\infty} \xi^2 p(\xi)\, d\xi = C < \infty.$$

Without loss of generality it can be assumed that g is a monotone estimate (i.e., $x_1 < x_2$ implies $g(x_1) \leqq g(x_2)$). Since the monotone estimates constitute a complete class (cf. [3]), any estimate which improves upon δ and is not monotone is in turn improved upon by a monotone estimate.

We add for the purposes of convenience the following assumption, which is so exceptionally weak as not to constitute any real restriction.

ASSUMPTION II. There exist constants a_1, a_2, b such that $a_1 < a_2$, $a_2 - a_1 < 1$, $b > 0$, and $p(\xi) \geqq b$ for $a_1 \leqq \xi \leqq a_2$.

Suppose there exists a sequence $\{x_i\}$ for which $g(x_i) - x_i \to \infty$ as $i \to \infty$. Then, there must exist for any n an index i_n such that $g(x_{i_n}) - x_{i_n} \geqq n$. Since g is monotone, $g(\xi) - \xi \geqq n - 1$ for $x_{i_n} \leqq \xi \leqq x_{i_n} + 1$. Let

$$\bar{\omega} = (a_1 + a_2)/2 - (x_{i_n} + \tfrac{1}{2}).$$

Then,

$$b(a_2 - a_1)(n - 1)^2 \leqq \int_{-\infty}^{\infty} (g(x) - x)^2 p(x + \bar{\omega})\, dx. \tag{22}$$

But by Lemma 2 the integral is bounded by $C < \infty$. Since n is arbitrary, this leads to a contradiction.

A similar argument applies if there exists a subsequence $\{x_i\}$ such that

$$g(x_i) - x_i \to -\infty$$

as $i \to \infty$. Thus, $g(x) - x$ must remain bounded and the admissibility of

$$\delta(x) = x$$

then follows according to Theorem 4 in the case where $p(x + \omega)$ has a monotone likelihood ratio.

The preceding argument also shows that in order for $|g(x) - x|$ to be unbounded and consistent with the result of Lemma 2 it must peak up very sparsely for durations of increasingly shorter lengths. Such pathologies are not excluded readily by means of our methods except for the two cases discussed. It seems unreasonable to admit such estimates for consideration.

A third case for which Theorem 4 is valid without the assumption of boundedness being necessary corresponds to the situation where $p(\xi)$ tends to zero exceptionally fast. More precisely, we assume that

$$\frac{\left| \int_{-\infty}^{u} \xi p(\xi)\, d\xi \right|}{p(u)} \leqq C. \tag{23}$$

For example this is satisfied by the standard normal distribution. The boundedness assumption was only used to prove Lemma 1 which on closer inspection is also valid whenever we can show that the expressions

$$\int_{-\infty}^{\infty} |g(x) - x|\, [-\Phi(x + n)]\, dx = A(n)$$

and

$$\int_{-\infty}^{\infty} |g(x) - x|\, [-\Phi(x - n)]\, dx = B(n)$$

are uniformly bounded. We study only the case of $A(n)$, the argument being

similar for $B(n)$. Invoking Schwarz's inequality, we obtain

$$A(n) \leqq \sqrt{\int [g(x) - x]^2 (-\Phi(x + n))\, dx} \sqrt{\int (-\Phi(u))\, du}$$

$$\leqq C\left\{\int [g(x) - x]^2 p(x + n)\, dx\right\}^{\frac{1}{2}} \leqq C',$$

where the last inequality is valid because of Lemma 2.

What we have shown for the general problem is that $\delta(x) = x$ is admissible within the class of all estimates g satisfying $|g(x) - x| \leqq M$ for all x where M is any finite constant. This means that x is admissible with respect to all estimates which do not differ too wildly from it. An appropriate formulation of the conclusions may be made in terms of a concept of local admissibility.

We close this section with a brief discussion of the case where the observation is integer-valued and the parameter ω also traverses the set of all integers. The analysis is considerably simpler.

In this case we can deduce immediately from the analog of Lemma 2 and equation (22) that if $\rho(\omega, g) \leqq \rho(\omega, \delta)$ for all integral ω then $|g(x) - x|$ is uniformly bounded.

Assumption I may be slightly weakened and now takes the form:

$$(24) \qquad \sum_x x^2 p(x) < \infty \qquad \text{and} \qquad \sum x\, p(x) = 0.$$

The role of Fourier transforms in the analysis is now replaced by Fourier series and the general line of the arguments carries over to the discrete case *mutatis mutandis*. Summing up we get:

THEOREM 5. *Suppose x and ω are discrete and integer-valued, and the conditions (24) are satisfied. If $\rho(\omega, g) \leqq \rho(\omega, \delta)$ where $\delta(x) = x$, then $g(x) \equiv x$, a.e.*

4. Translation parameter problem for the loss function $L(a, \omega) = (a + \omega)^{2N}$ with one observation. In the preceding section $\delta(x) = x$ was seen to be an admissible estimate of $-\omega$ in the translation parameter problem when the loss function is $L(a, \omega) = (a + \omega)^2$. Under suitable assumptions which are analogous to Assumption I and the boundedness restriction it will now be shown that $\delta(x) = x$ is also admissible for the loss function $L(a, \omega) = (a + \omega)^{2N}$. Note that consideration is still restricted to the case of a single observation.

The assumptions imposed are the following:

ASSUMPTION III. $p(\xi)$ is symmetric, i.e., $p(\xi) = p(-\xi)$.

Because of this assumption the odd moments of $p(\xi)$ vanish. This property is used in a crucial way.

If $h(x)$ is an estimate which presumably improves on $\delta(x)$, then we shall assume

ASSUMPTION IV. There exists a constant $M > 1$ such that $|h(x) - x| \leqq M$ for all x.

Several remarks will be appended pertaining to this assumption after the completion of the theorem. For suitable general classes of densities p we will find as before that Assumption IV is unnecessary.

ASSUMPTION V. $\int_{-\infty}^{\infty} \xi^{2N} p(\xi)\, d\xi < \infty$, $\int_{-\infty}^{\infty} \xi^{2N} p^2(\xi)\, d\xi < \infty$.

It is readily verified that the analogue of equation (18) is the following:

$$\begin{aligned} &\sum_{k=1}^{N} \binom{2N}{2k} \int_{-\infty}^{\infty} (x - h(x))^{2k} (x + \omega)^{2N-2k} p(x + \omega)\, dx \\ &\leq \sum_{k=0}^{N-1} \binom{2N}{2k+1} \int_{-\infty}^{\infty} (x - h(x))^{2k+1} (x + \omega)^{2N-2k-1} p(x + \omega)\, dx. \end{aligned} \tag{25}$$

The proofs of this section are completely analogous to those of the preceding section. Consequently, the detailed proofs will be shortened appropriately.

LEMMA 3. *If* $\rho(\omega, h) \leq \rho(\omega, \delta)$ *for all* ω, *then* $\int_{-\infty}^{\infty} | x - h(x)|^{\alpha}\, dx < \infty$ *for* $\alpha \geq 2$, (*under Assumptions* III, IV, *and* V).

PROOF. By (25), Fubini's theorem, and Assumption IV

$$\begin{aligned} &\sum_{k=1}^{N} \binom{2N}{2k} \int_{-\infty}^{\infty} (x - h(x))^{2k} \left\{ \int_{-n}^{n} (x + \omega)^{2N-2k} p(x + \omega)\, d\omega \right\} dx \\ &\leq M^{2N-1} \sum_{k=0}^{N-1} \binom{2N}{2k+1} \int_{-\infty}^{\infty} \left| \Phi_k(x + n) - \Phi_k(x - n) \right| dx = K(n), \end{aligned}$$

where $\Phi_k(u) = \int_{-\infty}^{u} \xi^{2N-2k-1} p(\xi)\, d\xi$. By Assumptions III and V it follows that $\Phi_k(u) \in L^1$. Thus

$$\begin{aligned} K(n) \leq M^{2N-1} \sum_{k=0}^{N-1} \binom{2N}{2k+1} \Bigg[&\int_{-\infty}^{\infty} | \Phi_k(x + n) |\, dx \\ &+ \int_{-\infty}^{\infty} | \Phi_k(x - n) |\, dx \Bigg] < C < \infty, \end{aligned}$$

where C is a constant independent n. Therefore,

$$0 \leq \sum_{k=1}^{N} \binom{2N}{2k} \left[\int_{-\infty}^{\infty} \xi^{2N-2k} p(\xi)\, d\xi \right] \left[\int_{-\infty}^{\infty} (x - h(x))^{2k}\, dx \right] < C.$$

THEOREM 6. *Let Assumptions* III, IV, *and* V *be satisfied.* $\rho(\omega, h) \leq \rho(\omega, \delta)$ *for all* ω *implies that* $h(x) = \delta(x)$, *a e.*

PROOF. The proof is obtained by adapting appropriately the methods employed in the discussion of Theorem 4. The details are omitted.

A few remarks promised earlier concerning Assumption IV will now be given. As in Section 3 for the case

$$p(\xi) = \begin{cases} 0, & \xi < a, \xi > b, -\infty < a < b < \infty, \\ \geq 0, & \text{otherwise}, \end{cases}$$

the only type of estimate which need be considered is an estimate $h(x)$ satisfying Assumption IV. The proof is the same as before. The argument for the second case in which $p(x, \omega) = p(x + \omega)$ has a monotone likelihood ratio is almost the same. It depends on Lemma 4 which may be derived with the aid of the Hölder inequality.

LEMMA 4. *If $\rho(\omega, h) \leqq \rho(\omega, \delta)$ for all ω, then there exists a constant C such that*

$$\int_{-\infty}^{\infty} (x - h(x))^{2N} p(x + \omega)\, dx \leqq C$$

for all ω (under Assumption V).

5. Translation parameter problem: n observations. The problem studied in this section is the multi-observation analogue of the problem treated in Section 3. Let $x_1, \cdots, x_n$ be n independent observations on the random variable X where X is distributed according to the density function $p(x, \omega) = p(x + \omega)$, $\omega \varepsilon (-\infty, \infty)$, with $p(\xi)$ a known prescribed density. Alternatively, X is allowed to be an integer-valued random variable with ω likewise assuming only integer values. $P\{X = i \mid \omega\} = p(i + \omega)$ where the probabilities $p(j)$, $j = 0, \pm 1, \pm 2, \cdots$, are assumed known. As previously the location parameter $-\omega$ is to be estimated.

Define $y_i = x_i - x_1$, $i = 2, \cdots, n$. An appealing estimate for the parameter $-\omega$ which was proposed by Pitman and has the property of being invariant with respect to translations of the observations x_i is

$$\delta^*(x_1, x_2, \cdots, x_n) = x_1 - T(y_2, \cdots, y_n), \tag{26}$$

where

$$T(y_2, \cdots, y_n) = \frac{\int_{-\infty}^{\infty} \xi p(\xi) \prod_{i=2}^{n} p(y_i + \xi)\, d\xi}{\int_{-\infty}^{\infty} p(\xi) \prod_{i=2}^{n} p(y_i + \xi)\, d\xi}. \tag{27}$$

Invariance of δ^* means that

$$\delta^*(x_1 + c, x_2 + c, \cdots, x_n + c) = c + \delta^*(x_1, x_2, \cdots, x_n)$$

for each constant c, an obviously desirable property when dealing with an unknown location parameter. It is well-known that δ^* is an invariant minimax estimator of $-\omega$ (cf. [5]).

Girshick and Savage [5] in discussing estimating procedures associated with quadratic loss conjectured that the estimator (26) is unique minimax. Since the risk of the estimate δ^* is identically constant it follows that in order to substantiate this conjecture it is enough to show that δ^* is admissible. This has been verified by Blackwell for the special case where both X and ω are essentially integer-valued and where $p(i)$ vanished except for at most a finite number of i [7]. He also constructed an example in which X traversed a discrete set and the range of ω was also discrete with values incommensurate with the possible X values, and he showed that δ^* need not be admissible in this case. This is not at all surprising in view of the fact that the usual demands corresponding to invariance in essence necessitate that the possible values of X and the ω values should comprise the same group structure. This characteristic was violated in the example of Blackwell.

We shall establish the admissibility of δ^* as an estimate of $-\omega$ in three separate cases which include most of the common distributions. In two of the cases we deal with densities of a continuous real variable for which ω traverses the real line. The third case examined is the general discrete problem where X and ω range over the integers. Blackwell's result for discrete densities with bounded domain emerges as a special case.

The convolution character of the location parameter problem suggests a representation of the problem in terms of Fourier integrals. It is therefore natural for our arguments to appeal to the powerful developed techniques of Fourier analysis which we, in fact, use abundantly. Our methods consequently apply to a considerably wider class of distributions which includes most of the common situations. The sequence of lemmas established follows principally the line of reasoning of the analogous single observation case and may be considered an extension thereof.

The three cases require separate analysis because of the different regularity assumptions needed for each. To establish the admissibility of $\delta^*(x, y_2, \cdots, y_n)$ we must show that if the inequality

$$
\begin{aligned}
\rho(\omega, g) = \int \cdots \int [g(x, y_2, \cdots, y_n) + \omega]^2 \\
\times p(x + \omega)p(x + \omega + y_2) \cdots p(x + \omega + y_n)\, dx dy_2 \cdots dy_n \\
\leq \int \cdots \int [\delta^*(x, y_2, \cdots, y_n) + \omega]^2 \\
\times p(x + \omega)p(x + \omega + y_2) \cdots p(x + \omega + y_n)\, dx dy_2 \cdots dy_n \\
= \rho(\omega, \delta^*) = c
\end{aligned} \tag{28}
$$

is valid for each ω then $g = \delta^*$, a.e. For the discrete case (i.e., X and ω are in teger-valued) the integral is to be replaced by the appropriate summation. The inequality (29) below is equivalent to (28).

$$
\begin{aligned}
\int \cdots \int [g(x, y_2, \cdots, y_n) - \delta^*(x, y_2, \cdots, y_n)]^2 \\
\times p(x + \omega) \prod_{i=2}^{n} p(x + \omega + y_i)\, dx dy_2 \cdots dy_n \\
\leq 2 \int \cdots \int [\delta^* - g][\delta^* + \omega] p(x + \omega) \prod_{i=2}^{n} p(x + \omega + y_i)\, dx dy_2 \cdots dy_n .
\end{aligned} \tag{29}
$$

CASE 1.

$$
p(\xi) \begin{cases} \geq 0, & -\infty < a \leq \xi \leq b < \infty, \\ = 0, & \xi < a, b < \xi. \end{cases}
$$

Only estimators $g(x_1, y_2, \cdots, y_n)$ of $-\omega$ which satisfy

$$
| \delta^*(x_1, y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)| \leq M
$$

for some constant $M < \infty$ need be considered. The argument is analogous to that given for the one-dimensional case in Section 3. The underlying reason is that the boundedness of the spectrum determines for each set of observations $x_1, x_2, \cdots, x_n$, an interval within which the true value of $-\omega$ must lie and any estimator which produces a value outside this interval can be improved upon. Thus any estimate which differs from δ^* and which improves in terms of risk on δ^* must only differ by a fixed constant from δ^*, regardless of the observed values of x. The single regularity assumption required in this case is that for all ξ, $0 \leqq p(\xi) \leqq C < \infty$.

LEMMA 5. *If* $\rho(\omega, g) \leqq \rho(\omega, \delta^*)$ *for all* ω, *then*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\delta^*(x_1, y_2, \cdots, y_n) - g(x_1, y_2, \cdots y_n)]^2 \times \left[\int_{-\infty}^{\infty} p(\xi)p(y_2 + \xi) \cdots p(y_n + \xi) d\xi\right] dx_1 dy_2 \cdots dy_n < \infty. \tag{30}$$

PROOF. By the fundamental inequality (29)

$$\begin{aligned} &\int_{-n}^{n}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\delta^*(x_1, y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)]^2 \\ &\quad\times p(x_1 + \omega) \cdots p(y_n + x_1 + \omega)\, dx_1 \cdots dy_n\, d\omega \\ &\leqq 2 \int_{-n}^{n}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\delta^*(x_1, y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)] \\ &\quad\times [\delta^*(x_1, y_2, \cdots, y_n) + \omega]p(x_1 + \omega) \cdots p(y_n + x_1 + \omega)\, dx_1 dy_2 \cdots dy_n\, d\omega \\ &\leqq 2M \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} | \Phi(x_1 + n, y_2, \cdots, y_n) \\ &\qquad - \Phi(x_1 - n, y_2, \cdots, y_n) |\, dx_1 dy_2 \cdots dy_n, \end{aligned} \tag{31}$$

where

$$\Phi(u, y_2, \cdots, y_n) = \int_{-\infty}^{u} [\xi - T(y_2, \cdots, y_n)]p(\xi)p(y_2 + \xi) \cdots p(y_n + \xi)\, d\xi.$$

By direct calculation we observe that

$$\Phi(\infty, y_2, \cdots, y_n) = \Phi(-\infty, y_2, \cdots, y_n) = 0. \tag{32}$$

If we can show that $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |\Phi(x_1, y_2, \cdots, y_n)|\, dx_1 dy_2 \cdots dy_n < \infty$, then it follows that the expression of (31) is uniformly bounded with respect to n which clearly implies the sought-for conclusion. The remainder of the proof consists in verifying the finiteness of this integral.

Note that $\Phi(x_1, y_2, \cdots, y_n) \leqq 0$ for all $x_1, y_2, \cdots, y_n$, and for fixed $y_2, \cdots, y_n$ there exists a constant N such that $|x_1| \geqq N$ implies

$$\Phi(x_1, y_2, \cdots, y_n) = 0.$$

Integration by parts with respect to x_1 yields

$$0 \leqq - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi(x_1, y_2, \cdots, y_n)\, dx_1 dy_2 \cdots dy_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1[x_1 - T(y_2, \cdots, y_n)]p(x_1) \cdots p(y_n + x_1)\, dx_1 \cdots dy_n .$$

But the last integral converges absolutely since $p(\xi)$ vanishes outside a finite interval and $| T(y_2, \cdots, y_n)| \leqq | a | + | b |$.

THEOREM 7. *If* $\rho(\omega, g) \leqq \rho(\omega, \delta^*)$ *for all* ω, *then* $g = \delta^*$, *a.e.*

PROOF. Let

$$G(x_1, y_2, \cdots, y_n) = [x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)] \times \left[\int_{-\infty}^{\infty} p(\xi) \cdots p(y_n + \xi)\, d\xi \right]^{1/2},$$

$$H(u, y_2, \cdots, y_n) = [u - T(y_2, \cdots, y_n)] \left[\int_{-\infty}^{\infty} p(\xi) \cdots p(y_n + \xi)\, d\xi \right]^{-1/2} \times [p(u)p(y_2 + u) \cdots p(y_n + u)].$$

By Lemma 5, $G(x_1, y_2, \cdots, y_n) \varepsilon L^2$, and by direct calculation we see that $H(u, y_2, \cdots, y_n) \varepsilon L^2$. Therefore, the Fourier transforms $\tilde{G}(t_1, \cdots, t_n)$ and $\tilde{H}(t_1, \cdots, t_n)$ of G and H, respectively, are well-defined and belong to L^2. Consider the expression

$$\begin{aligned} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} & [x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)] \\ & \times [x_1 + \omega_1 - T(y_2 + \omega_2, \cdots, y_n + \omega_n)] \\ & \times p(x_1 + \omega_1) \cdots p(y_n + \omega_n + x_1 + \omega_1)\, dx_1 dy_2 \cdots dy_n \end{aligned} \tag{33}$$

where for our purposes we shall need to evaluate this expression only for the values $\omega_2 = \cdots = \omega_n = 0$. (33) is essentially a convolution of G and H so its Fourier transform exists and is equal to

$$\tilde{G}(-t_1, -t_2, \cdots, -t_n)\tilde{H}(t_1, t_2, \cdots, \text{[illegible]}$$

By the inversion property of Fourier transforms we [illegible]

$$\begin{aligned} \int_{-n}^{n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} & [x_1 - T(y_2, \cdots, y_n) - g(\text{[illegible]})] \\ & \times [x_1 + \omega_1 - T(y_2, \cdots, y_n)] \\ & \times p(x_1 + \omega_1) \cdots p(y_n + x_1 \text{[illegible]}) \cdots d\omega_1 \end{aligned} \tag{34}$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\tilde{G}(-t_1, \cdots, -t_n)\tilde{H}(\text{[illegible]})}{\text{[illegible]}} \text{[illegible]}$$

which is defined everywhere since $\tilde{G}\cdot\tilde{H}$ belongs to L^1. The last integral is an absolutely convergent integral. To substantiate this assertion we note that $\tilde{H}(0, t_2, \cdots, t_n) = 0$, and

$$\frac{\partial}{\partial t_1}\tilde{H}(t_1, \cdots, t_n)$$
$$= i\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[i(t_1x_1 + \cdots + t_nx_n)\right]x_1[x_1 - T(y_2, \cdots, y_n)]$$
$$\times\left[\int_{-\infty}^{\infty}p(\xi)\cdots p(y_n+\xi)\,d\xi\right]^{-1/2}p(x_1)p(y_2+x_1)\cdots p(y_n+x_1)\,dx_1\cdots dy_n$$

is bounded independently of $t_1, \cdots, t_n$. Hence, by the mean value theorem $\tilde{G}(-t_1, \cdots, -t_n)\tilde{H}(t_1, \cdots, t_n)/t_1$ as a function of t_1 is integrable in a neighborhood about the origin for all $t_2, \cdots, t_n$. Therefore,

$$\tilde{G}(-t_1, \cdots, -t_n)\tilde{H}(t_1, \cdots, t_n)/t_1 \in L^1.$$

By virtue of the Riemann-Lebesgue lemma and Lebesgue's theorem of dominated convergence we see that the expression in (34) tends to zero as $n \to \infty$. Hence, on comparison of (29) and (34), we deduce

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots y_n)]^2$$
$$\times\left[\int_{-\infty}^{\infty}p(x_1+\omega)\cdots p(y_n+x_1+\omega)\,d\omega\right]dx_1\cdots dy_n \leqq 0,$$

which establishes the theorem.

CASE 2. Discrete case.

$$X = \{0, \pm1, \pm2, \cdots\}, \qquad \Omega = \{0, \pm1, \pm2, \cdots\},$$

and

$$P\{x = i \mid \omega\} = p(i+\omega),$$

where $p(j) \geqq 0$, $\sum_{j=-\infty}^{\infty} p(j) = 1$.

We impose the following regularity assumption.

ASSUMPTION VI. $\sum_{j=-\infty}^{\infty} j^2\sqrt{p(j)} < \infty$.

Unfortunately, we do not know whether this assumption may be relaxed to the obviously weaker and more natural condition $\sum_{-\infty}^{\infty} j^2p(j) < \infty$. The weaker requirement was indeed sufficient for the case of a single observation whenever $\sum_{-\infty}^{\infty} jp(j) = 0$. [See Section 3.]

LEMMA 6. *If $\rho(\omega, g) \leqq \rho(\omega, \delta^*)$ for all ω, then there exists a constant C such that*

$$\sum_{y_n}\cdots\sum_{x_1}[\delta^*(x_1, y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)]^2$$
$$\times p(x_1+\omega)\cdots p(y_n+x_1+\omega) \leqq C$$

for all ω (under Assumption VI).

The proof is analogous to that of Lemma 2 of Section 3 so it is omitted.

LEMMA 7. *If $\rho(\omega, g) \leqq \rho(\omega, \delta^*)$ for all ω, then*

$$\sum_{y_n} \cdots \sum_{x_1} [\delta^*(x_1, y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)]^2 \times [\sum_{\omega} p(x_1 + \omega) \cdots p(y_n + x_1 + \omega)] < \infty.$$

PROOF. As a consequence of Lemma 6,

$$| x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n) | \leqq \frac{C^{1/2}}{\sqrt{p(x_1 + \omega) \cdots p(y_n + x_1 + \omega)}}$$

for all ω. Since ω is arbitrary,

$$\max_{x_1} | x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n) | \leqq \frac{C^{1/2}}{\max_j \sqrt{p(j)p(y_1 + j) \cdots p(y_n + j)}}.$$

Define for integers u,

$$\Phi(u, y_2, \cdots, y_n) = \sum_{j=-\infty}^{u} [j - T(y_2, \cdots, y_n)]p(j)p(y_2 + j) \cdots p(y_n + j).$$

By the fundamental inequality (29)

$$\begin{aligned}\sum_{\omega=-n}^{n} \sum_{y_n} \cdots \sum_{x_1} & [x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)]^2 \\ & \times p(x_1 + \omega) \cdots p(y_n + x_1 + \omega) \\ & \leqq 2 \sum_{y_n} \cdots \sum_{x_1} \frac{C^{1/2}}{\max_j \sqrt{p(j)p(y_2 + j) \cdots p(y_n + j)}} \\ & \times | \Phi(x_1 + n, y_2, \cdots, y_n) - \Phi(x_1 - n, y_2, \cdots, y_n) |.\end{aligned} \tag{35}$$

It is easily checked that $\Phi(u, y_2, \cdots, y_n) \leqq 0$ for all $u, y_2, \cdots, y_n$, and as $| u | \to \infty$ $| u\Phi(u, y_2, \cdots, y_n)| \to 0$ by Assumption VI with the aid of the fact that $\Phi(\infty, y_2, \cdots, y_n) = 0$. Summation by parts with respect to x_1 yields

$$\begin{aligned}-\sum_{y_n} \cdots \sum_{x_1} & \frac{\Phi(x_1, y_2, \cdots, y_n)}{\max_j \sqrt{p(j)p(y_2 + j) \cdots p(y_n + j)}} \\ & = \sum_{y_n} \cdots \sum_{x_1} \frac{x_1[x_1 - T(y_2, \cdots, y_n)]p(x_1)p(y_2 + x_1) \cdots p(y_n + x_1)}{\max_j \sqrt{p(j)p(y_2 + j) \cdots p(y_n + j)}} \\ & \leqq 2 \sum_{y_n} \cdots \sum_{x_1} \frac{x_1^2 p(x_1)p(y_2 + x_1) \cdots p(y_n + x_1)}{\max_j \sqrt{p(j)p(y_2 + j) \cdots p(y_n + j)}} \\ & \leqq 2 \sum_{y_n}{}' \cdots \sum_{x_1}{}' \frac{x_1^2 p(x_1) \cdots p(y_n + x_1)}{\sqrt{p(x_1) \cdots p(y_n + x_1)}},\end{aligned} \tag{36}$$

where $\sum'_{y_n} \cdots \sum'_{x_1}$ denotes summation over all $x_1, \cdots, y_n$ for which

$$p(x_1)p(y_2 + x_1) \cdots p(y_n + x_1) > 0.$$

But by Assumption VI

$$\sum_{y_n} \cdots \sum_{x_1} x_1^2 \sqrt{p(x_1) \cdots p(y_n + x_1)} < \infty. \tag{37}$$

Hence, (35), (36), and (37) in conjunction yield the desired result.

THEOREM 8. *If* $\rho(\omega, g) \leqq \rho(\omega, \delta^*)$ *for all* ω, *then*

$$g(x_1, y_2, \cdots, y_n) = \delta^*(x_1, y_2, \cdots, y_n)$$

for all $y_2, \cdots, y_n$ *such that* $\sum_j p(j)p(y_2 + j) \cdots p(y_n + j) > 0$.

PROOF. Let

$$G(x_1, y_2, \cdots, y_n) = [x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)] \times [\sum_j p(j) \cdots p(y_n + j)]^{1/2}$$

$$H(u, y_2, \cdots, y_n) = [u - T(y_2, \cdots, y_n)] [\sum_j p(j)p(y_2 + j) \cdots p(y_n + j)]^{-1/2} \times [p(u)p(y_2 + u) \cdots p(y_n + u)]$$

Since $\sum_{y_n} \cdots \sum_{x_1} G^2(x_1, y_2, \cdots, y_n) < \infty$ by Lemma 7,

$$\sum_{y_n} \cdots \sum_{x_1} \exp [i(t_1 x_1 + \cdots + t_n y_n)]G(x_1, y_2, \cdots, y_n)$$

converges in quadratic mean to a function $\tilde{G}(t_1, \cdots, t_n) \varepsilon L^2(-\pi, \pi)$. Also, b[y] Assumption VI $\sum_{y_n} \cdots \sum_{x_1} | H(x_1, y_2, \cdots, y_n)| < \infty$. Indeed, inspection [of] the series shows that its convergence would be a consequence of the conve[r]gence of the related series

$$\sum_{u, y_2, \cdots, y_n} \frac{| u | p(u)p(u + y_2) \cdots p(u + y_n)}{\sqrt{\sum_\xi p(\xi)p(\xi + y_2) \cdots p(\xi + y_n)}} = J.$$

This follows in view of the inequality of Schwarz, the uniform boundedness [of]

$$\frac{p(u)p(u + y_2) \cdots p(u + y_n)}{\sum_\xi p(\xi)p(\xi + y_2) \cdots p(\xi + y_n)},$$

and Assumption VI. Hence,

$$\sum_{y_n} \cdots \sum_{x_1} \exp [i(t_1 x_1 + \cdots + t_n y_n)]H(x_1, y_2, \cdots, y_n)$$

converges uniformly and absolutely to a function $\tilde{H}(t_1, \cdots, t_n) \varepsilon L^2(-\pi, \pi$

The expression

$$I(\omega_1, \cdots, \omega_n) = \sum_{y_n} \cdots \sum_{x_1} [x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)] \times [x_1 + \omega_1 - T(y_2 + \omega_2, \cdots, y_n + \omega_n)] \times p(x_1 + \omega_1) \cdots p(y_n + \omega_n + x_1 + \omega_1),$$

where in actuality $\omega_2 = \cdots = \omega_n = 0$, is essentially a convolution of G and H so its Fourier series converges absolutely to a function $\bar{I}(t_1, \cdots, t_n)$, and $\bar{I}(t_1, \cdots, t_n) = \tilde{G}(-t_1, \cdots, -t_n)\tilde{H}(t_1, \cdots, t_n)$, a.e. Since

$$I(\omega_1, 0, \cdots, 0) = 1/(2\pi)^n \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \exp(-it_1\omega_1)\bar{I}(t_1, \cdots, t_n)\, dt_1 \cdots dt_n$$

and $\sum_{-n}^{n} e^{-itk} = [e^{itn} - e^{-it(n+1)}]/(1 - e^{-it})$, it follows that

$$\text{(38)} \quad \sum_{\omega_1=-n}^{n} I(\omega_1, 0, \cdots, 0) = \frac{1}{(2\pi)^n} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \bar{I}(t_1, \cdots, t_n) \times \left[\frac{e^{it_1 n} - e^{-it_1(n+1)}}{1 - e^{-it_1}}\right] dt_1 \cdots dt_n .$$

The interchange of summation and integration signs on the right-hand side of (38) is valid since by virtue of Assumption VI

$$\lim_{t_1 \to 0} \bar{I}(t_1, \cdots, t_n)/(1 - e^{-it_1}) < \infty$$

and $\bar{I}(t_1, \cdots, t_n) \in L^1(-\pi, \pi)$. But by the Riemann-Lebesgue lemma the right-hand side of (38) converges to zero as $n \to \infty$. By the fundamental inequality

$$\sum_{y_n} \cdots \sum_{x_1} [x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)]^2 [\sum_j p(j) \cdots p(y_n + j)] \leqq 0$$

from which the desired result follows

Case 3. General density functions.

This case will include all density functions which satisfy the following regularity conditions:

Assumption VII:

$$0 \leqq p(\xi) \leqq C, \int_{-\infty}^{\infty} \xi^2 \sqrt{p(\xi)}\, d\xi < \infty.$$

Assumption VIII:

$$\frac{p(\xi)p(y_2 + \xi) \cdots p(y_n + \xi)}{\int_{-\infty}^{\infty} p(\theta)p(y_2 + \theta) \cdots p(y_n + \theta)\, d\theta} \leqq C' \quad \text{for all } \xi, y_2, \cdots, y_n .$$

Assumption VIII asserts that the conditional density of x_1 given $y_2, \cdots, y_n$ must remain bounded for all $x_1, y_2, \cdots, y_n$. This assumption is a bit stronger

than necessary; it could be replaced by an assumption of finiteness of a number of definite integrals involving the conditional density. However, there seems to be no gain involved in such a generalization. The class of densities which satisfy Assumptions VII and VIII includes as two of its important members the normal and negative exponential distributions as well as any density which asymptotically dies off like a power.

It will be shown by Theorem 9 below that δ^* is admissible with respect to the class of all estimators $g(x_1, y_2, \cdots, y_n)$ which satisfy the following additional requirement:

ASSUMPTION IX. There exists a constant $M < \infty$ such that for all $x_1, y_2, \cdots, y_n$ $|\delta^*(x_1, y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)| \leq M$.

This will establish a suitably broad form of the concept of "local" admissibility for the estimator δ^*. This concept of "local" admissibility was introduced earlier in Section 3. As yet suitable supplementary conditions on the form of $p(\xi)$ for the relaxation of this assumption have not been obtained.

LEMMA 8. *If $\rho(\omega, g) \leq \rho(\omega, \delta^*)$ for all ω, and g satisfies Assumption* IX, *then*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\delta^*(x_1, y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)]^2$$
$$\times \left[\int_{-\infty}^{\infty} p(\xi)p(y_2 + \xi) \cdots p(y_n + \xi)\, d\xi\right] dx_1 dy_2 \cdots dy_n < \infty.$$

PROOF. The proof is analogous to that of Lemma 5. It is sufficient to prove that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |\Phi(x_1, y_2, \cdots, y_n)|\, dx_1 dy_2 \cdots dy_n < \infty,$$

where Φ is defined as in Lemma 5. $\Phi(x_1, y_2, \cdots, y_n) \leq 0$, and as $|u| \to \infty$, $|u\Phi(u, y_2, \cdots, y_n)| \to 0$. As $u \to -\infty$

$$\begin{aligned} 0 &\leq u\,\Phi(u, y_2, \cdots, y_n) \\ &\leq \int_{-\infty}^{u} \xi^2 p(\xi)p(y_2 + \xi) \cdots p(y_n + \xi)\, d\xi \\ &\quad + |T(y_2, \cdots, y_n)| \int_{-\infty}^{u} \xi p(\xi)p(y_2 + \xi) \cdots p(y_n + \xi)\, d\xi. \end{aligned} \tag{39}$$

Both integrals in (39) vanish as $u \to -\infty$ by Assumption VII. A similar analysis is valid as $u \to \infty$ since

$$\int_{-\infty}^{u} [\xi - T(y_2, \cdots, y_n)]p(\xi) \cdots p(y_n + \xi)\, d\xi$$
$$= -\int_{u}^{\infty} [\xi - T(y_2, \cdots, y_n)]\, p(\xi) \cdots p(y_n + \xi)\, d\xi.$$

Integration by parts with respect to x_1 and an application of Schwarz's inequality yields

$$0 \leqq - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi(x_1, y_2, \cdots, y_n)\, dx_1\, dy_2 \cdots dy_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1 [x_1 - T(y_2, \cdots, y_n)]\, p(x_1) \cdots p(y_n + x_1)\, dx_1 \cdots dy_n$$

$$\leqq 2 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^2\, p(x_1) \cdots p(y_n + x_1)\, dx_1 \cdots dy_n .$$

The final expression is therefore finite by virtue of Assumption VII.

THEOREM 9. *If $\rho(\omega, g) \leqq \rho(\omega, \delta^*)$ for all ω, and g satisfies Assumption* IX, *then $g = \delta^*$, a.e.*

PROOF. Define $G(x_1, y_2, \cdots, y_n)$ and $H(x_1, y_2, \cdots, y_n)$ as in Theorem 7. By Lemma 8, $G(x_1, y_2, \cdots, y_n) \,\varepsilon\, L^2$, and by Assumptions VII and VIII, $H(x_1, y_2, \cdots, y_n) \,\varepsilon\, L^2$. Therefore, the Fourier transforms $\bar{G}(t_1, \cdots, t_n)$ and $\bar{H}(t_1, \cdots, t_n)$ of G and H, respectively, are well-defined and belong to L^2. The Fourier transform $\bar{I}(t_1, \cdots, t_n)$ of

$$I(\omega_1, \cdots, \omega_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [x_1 - T(y_2, \cdots, y_n) - g(x_1, y_2, \cdots, y_n)]$$
$$\times [x_1 + \omega_1 - T(y_2 + \omega_2, \cdots, y_n + \omega_n)]\, p(x_1 + \omega_1) \cdots p(y_n + \omega_n + x_1 + \omega_1)$$
$$\times dx_1 dy_2 \cdots dy_n$$

is well-defined and equals $\bar{G}(-t_1, \cdots, -t_n)\bar{H}(t_1, \cdots, t_n)$ By an argument analogous to that of Theorem 7 it follows that

$$\lim_{n \to \infty} \int_{-n}^{n} I(\omega_1, 0, \cdots, 0)\, d\omega_1 = 0,$$

which proves the result.

6. Acknowledgement. In closing, the author acknowledges many valuable discussions with C. Stein regarding the translation parameter problem. We are also informed that he has developed an alternative approach for estimating a translation parameter using the ideas of Bayes estimates.

REFERENCES

[1] DAVID BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, 1954.
[2] E. L. LEHMANN, "Notes on the theory of estimation," (mimeographed notes recorded by Colin Blyth), University of California, Berkeley, 1950.
[3] S. KARLIN AND H. RUBIN, "The theory of decision procedures for distributions with monotone likelihood ratio," *Ann Math Stat.*, Vol 27 (1956), pp. 272-299.

[4] J. L. Hodges, Jr. and E. L. Lehmann, "Some applications of the Cramér-Rao inequality," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 13-22.
[5] M. A. Girshick and L. J. Savage, "Bayes and Minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 53-73.
[6] David Blackwell, "Conditional expectation and unbiased sequential estimation," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 105-110.
[7] David Blackwell, "On the translation parameter problem for discrete variables," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 393-399.
[8] Colin R. Blyth, "On minimax statistical decision procedures and their admissibility," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 22-42.

ON THE ESTIMATION OF PARAMETERS RESTRICTED BY INEQUALITIES[1]

By H. D. Brunk

University of Missouri

1. Summary. There are collected in this paper several observations and results more or less loosely related by their connections with the subject mentioned in the title. The discussion moves from the general to the specific, beginning with some remarks on minimization of convex functions subject to side conditions, and ending with a discussion of uniform consistency of estimators of linearly ordered parameters.

Section 2 deals with one aspect of the problem of minimizing a function of several variables, subject to side conditions which specify that the variables must satisfy certain inequalities. It is frequently true in such problems that information as to which of the restricting sets contain the minimizing point on their boundaries is of great assistance in finding this point. Theorem 2.1 provides the basis for a stepwise procedure leading to this information when both the function to be minimized and the restricting sets are convex. It makes no contribution, however, to the problem of finding the minimizing point on a given boundary or intersection of boundaries

Brief mention is made in Section 3 of some examples of estimation problems for which the remark to which Section 2 is devoted is appropriate.

Section 4 is concerned with a situation in which samples are taken from k populations, each known to belong to a given one-parameter "exponential family". The problem is the maximum likelihood estimation of the k parameters determining the populations, subject to certain restrictions. Methods are discussed of finding the minimizing point on a given intersection of boundaries of restricting sets. In the particular case when all populations belong to the same exponential family and when the restrictions on the parameters are order restrictions, it is observed that the maximum likelihood estimators (MLE's) of the means are independent of the particular exponential family.

In Section 5 is discussed a property, related to sufficiency, of the MLE's discussed in Section 4. Let y denote a vector representing a set of possible values of the MLE's, E a Borel subset of the sample space, τ a parameter point, S_0 the intersection of the restricting sets. If S_0 is bounded by hyperplanes, there is a determination of the conditional probability $p_\tau(E \mid y)$ which is independent of τ when y is interior to S_0, and, when y lies on a face, edge, or vertex of S_0, is independent of τ on the closure of that face, edge, or vertex. This result may

Received September 10, 1956; revised December 16, 1957.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under contract No. AF 18(600)-1108.

be regarded as a generalization of a remark ([16], p. 77) to the effect that if X and Y are normally distributed random variables with unit standard deviation and means ξ and η respectively, and if ξ and η are known to satisfy a linear equation, then the foot of the perpendicular from the observation point (x, y) to the line is a sufficient estimator.

Section 6 is devoted to the same problem as are Sections 4 and 5, except that the parameters are linearly ordered, and that the populations need not belong to exponential families. Conditions are obtained for the strong uniform consistency of an estimator which is the MLE when the populations do belong to the same exponential family. An asymptotic lower bound is given for the probability of achieving a given precision uniformly.

2. Minimizing a convex function on the intersection of closed convex sets. (The author's thanks are due the referee, whose suggestions have materially improved the exposition in this section.) Let $y = (y_1, y_2, \cdots, y_k)$ denote the generic point of R_k, Euclidean space of k dimensions, and let $G(y)$ be a lower semi-continuous function such that $\{y : G(y) \leqq a\}$ is bounded for each a, satisfying

$$G[\lambda y' + (1 - \lambda)y''] \leqq \max [G(y'), G(y'')] \tag{2.1}$$

for $0 \leqq \lambda \leqq 1$, and for all y', y'' in its (convex) domain of definition. (This form of condition (2.1) is due to the referee.) In particular, G satisfies (2.1) if G is convex.

For an arbitrary set $A \subset R_k$, let $\mathfrak{B}(A)$ denote its boundary. We write $A \subset B$ if A is properly contained in B or if $A = B$. Let ϕ denote the empty set. Let there be given a finite number of intersecting closed convex sets A_i $(i = 1, 2, \cdots, N)$. We assume G defined on a convex set containing $\bigcup_{i=1}^N A_i$. We define Q_i to be the set on which $G(y)$ achieves its minimum value for $y \in A_i$, $i = 1, 2, \cdots, N$. For a set $i_1, i_2, \cdots, i_n$ of distinct positive integers not greater than N we define $Q_{i_1, i_2, \cdots, i_n}$ to be the set on which $G(y)$ achieves its minimum value for $y \in A_{i_1} A_{i_2} \cdots A_{i_n}$.

THEOREM 2.1. *Let A_1, A_2 be intersecting closed sets, A_1 convex. Then either $Q_{12} \subset Q_1$ or $Q_{12}\mathfrak{B}(A_2) \neq \phi$.*

PROOF. If $Q_1 A_2 \neq \phi$ then obviously $Q_1 \supset Q_{12}$. It remains to consider the situation in which $Q_1 A_2 = \phi$. Let $p \in Q_1$, $q \in Q_{12}$. Since A_1 is convex, the segment pq lies in A_1. Since $p \notin A_2$, $q \in A_2$, there is a point r on pq such that $r \in A_1 \mathfrak{B}(A_2)$. By property (2.1), $G(r) \leqq G(q)$, hence $r \in Q_{12}$. This completes the proof of Theorem 2.1.

COROLLARY 2.1. *If $G(y)$ is lower semi-continuous, if $\{y : G(y) \leqq a\}$ is bounded for each a and if G satisfies*

$$G[\lambda y' + (1 - \lambda)y''] < \max [G(y'), G(y'')] \tag{2.2}$$

for $0 < \lambda < 1$, and for all y', y'' in its (convex) domain of definition, and if A_1, A_2 are intersecting closed convex sets in its domain of definition, then Q_1 and Q_{12} consist of single points, q_1 and q_{12}; either $q_1 = q_{12}$ or $q_{12} \in \mathfrak{B}(A_2)$. We note that a strictly convex function G satisfies (2.2).

Corollary 2.1 justifies the procedure outlined in the following paragraph for minimizing G subject to the condition $y \varepsilon A_1A_2, \cdots, A_N$, where $A_1, A_2, \cdots, A_N$ are given intersecting closed convex sets. In many particular instances of this problem, one of the chief difficulties is that of determining which of the sets A_i contain the solution (a point minimizing G) on their boundaries, when the point at which G attains its unrestricted minimum is not in $A_1A_2, \cdots, A_N$. The procedure described below can be used to determine those sets among $A_1, A_2, \cdots, A_N$ on whose boundaries the solution lies. We remark that G need not be convex in order for the method to apply, provided it is lower semi-continuous and satisfies (2.2).

The first step is to determine the point at which G assumes its unrestricted minimum. If this point lies in $A_1A_2, \cdots, A_N$, it is the solution. If not, one of the sets is selected in which it does not lie, and designated as A_1 (relabelling, if necessary). Now consider the problem of minimizing G subject to $y \varepsilon A_1$. Applying Corollary 2.1, with A_1 there replaced by the whole space in this application, and A_2 there by A_1 in this application, we find that the solution, q_1, lies on $\mathcal{B}(A_1)$. It may be that q_1 lies in $A_1A_2, \cdots, A_N$, in which case it is the solution. If not, we designate as A_2 (relabelling, if necessary) one of the sets which does not contain q_1. We now consider the problem of minimizing G subject to $y \varepsilon A_1A_2$. By Corollary 2.1, the solution q_{12} lies on $\mathcal{B}(A_2)$. We find first the point q_2 where G is minimized subject to $y \varepsilon A_2$. If $q_2 \varepsilon A_1A_2$, then $q_2 = q_{12}$ is the solution of the present limited problem. Otherwise, by another application of Corollary 2.1, $q_{12} \varepsilon \mathcal{B}(A_1)\mathcal{B}(A_2)$, etc.

This stepwise procedure was introduced in situations involving certain functions G and convex sets A_i described by inequalities of the form $y_j \leqq y_k$ by van Eeden ([13], Theorem I, p. 445; [14], Theorem II, p. 134). The stepwise procedure outlined above makes no contribution to the problem of finding the point where G is minimized on a given "extended hyperface" $\mathcal{B}(A_{i_1}) \cdots \mathcal{B}(A_{i_n})$. Further, in special cases it may even occur that one will determine the minimizing point on each of the 2^N-1 "extended hyperfaces" before finding a minimizing point in $A_1A_2, \cdots, A_N$. Usually, however, one will expect the procedure to terminate with the solution long before all "extended hyperfaces" have been examined.

Non-linear programming methods have been developed for solving certain problems of this class (see, for example, [3]). Problems arising from some of the applications discussed below are such that it is relatively easy to find the minimizing point on a given "extended hyperface", and some trial calculations with such problems using the above stepwise procedure resulted in far less lengthy calculations than did those using general nonlinear programming methods.

3. Examples.

(i) In the bioassay type of problem, one is required to minimize a convex function of the form

$$-\sum_{i=1}^{N} [a_i \log y_i + b_i \log (1 - y_i)], \tag{3.1}$$

where the a_i and b_i are given numbers, and the y_i are subject to the restriction

$$0 \leqq y_1 \leqq y_2 \leqq \cdots \leqq y_N \leqq 1.$$

Even if one is not willing to assume a particular form for the distribution function and is thus led to this nonparametric formulation, he may feel that, for example, the distribution function should not rise too rapidly, and be led to impose further conditions of the form

$$y_{i+1} - y_i \leqq c_i \text{ or } y_{i+2} - 2y_{i+1} + y_i \leqq d_i \tag{3.2}$$

where the c_i and d_i are prescribed numbers. The problem remains in the class discussed in Section 2; however, the minimizing point on the boundary of a set described by inequalities of the form (3.2) is not in general so easily found as is that on a boundary $y_i = y_{i+1}$. The fact that the partial derivatives of the function (3.1) are so readily determined suggests that the method of Lagrange's multipliers, together with Newton's (multivariate) method for solution of simultaneous equations may prove appropriate.

A similar but simpler problem might conceivably arise in connection with ordinary random sampling. Let $x_1, \cdots, x_n$ be sample values of a sample of size n from a population with unknown distribution function F, and let $p_1, p_2, \cdots, p_n$ be the salti or jumps of F at the sample values. The MLE's of $p_1, p_2, \cdots, p_n$ maximize $\prod_{i=1}^n p_i$ or minimize $-\sum_{i=1}^n \log p_i$ subject to the restriction $\sum_{i=1}^n p_i = 1$, and are given by $p_i = 1/n$, $i = 1, 2, \cdots, n$, furnishing the empiric distribution function. But now if we suppose further conditions put on F, perhaps of the form $F(x_{i+1}) - F(x_i) \leqq c(x_{i+1} - x_i)$ or $p_i \leqq c(x_{i+1} - x_i)$, $i = 1, 2, \cdots, n - 1$, the remark of Section 2 may prove useful.

(ii) In the example on page 833 in [6], one is given $\{\alpha_{ij}\}$, $\{n_{ij}\}$, and required to choose $\{p_{ij}\}$ so as to minimize

$$-\sum_{i=1}^{n} \sum_{j=1}^{k} [\alpha_{ij} \log p_{ij} + (n_{ij} - \alpha_{ij}) \log (1 - p_{ij})].$$

Here $p_{ij} = 1 - F(x_i, y_j)$, where x_i, $i = 1, 2, \cdots, n$, and y_j, $j = 1, 2, \cdots, k$ are given, and where $F(x, y)$ is an unknown bivariate distribution function, so that not only is it required to be monotone in the two variables separately, but also second differences are to be positive.

(iii) Let a person chosen at random from a group have a probability U of contracting a certain disease in unit time; U is to be considered a random variable, with distribution function F. If a particular person has probability u_0, then the probability that he will be infected for the first time during a second unit of time is $(1 - u_0)u_0$, infected for the first time during a third is $(1 - u_0)^2u_0$, etc. Thus the probability that a person chosen at random will become infected during the first unit of time is

$$p_1 = \int_0^1 u \, dF(u) = \int_0^\infty (1 - e^{-t}) \, dG(t),$$

where $G(t) = F(1 - e^{-t})$; the probability that he will first become infected during the jth unit of time is $p_j = \int_0^1 (1 - u)^{j-1} u \, dF(u) = \int_0^\infty e^{-(j-1)t} (1 - e^{-t}) \, dG(t)$, $j = 1, 2, \cdots$. If we set $q_j = \int_0^\infty e^{-jt} \, dG(t), j = 0, 1, 2, \cdots$, then $p_j = q_{j-1} - q_j$, $j = 1, 2, \cdots$, and $q_j = 1 - \sum_{i=1}^{j} p_i$, $j = 1, 2, \cdots$. Since G is a distribution function, we have

$$\Delta_j q = q_{j+1} - q_j = -p_{j+1} \leqq 0,$$

$$\Delta_j^2 q = q_{j+2} - 2q_{j+1} + q_j = -(p_{j+2} - p_{j+1}) \geqq 0, \text{ etc.}$$

Suppose that of n persons initially chosen at random, x_j first become infected during the jth unit of time, $j = 1, 2, \cdots, k$, and that $x_{k+1} = n - \sum_{j=1}^{k} x_j$ fail to become infected during the first k units of time. The MLE's of the probabilities p_j $(j = 1, 2, \cdots, k)$ and $1 - \sum_{j=1}^{k} p_j$ are the solutions $y_1, y_2, \cdots, y_k, y_{k+1}$ of the following problem: to minimize

$$-\sum_{j=1}^{k+1} x_j \log y_j,$$

subject to

$$\sum_{j=1}^{k+1} y_j = 1, \tag{3.3}$$

and

$$\begin{cases} 0 \leqq y_j \leqq 1, j = 1, 2, \cdots, k+1, \\ y_{j+1} - y_j \leqq 0, j = 1, 2, \cdots, k-1, \\ y_{j+2} - 2y_{j+1} + y_j \geqq 0, j = 1, 2, \cdots, k-2, \text{ etc.} \end{cases} \tag{3.4}$$

The problem may be made to fit precisely the pattern of Section 2 if we replace (3.3) by

$$\sum_{j=1}^{k+1} y_j \leqq 1,$$

the altered problem clearly has the same solution.

4. Exponential families. The remark to which Section 2 is devoted is especially appropriate for the problem of estimating parameters using samples from populations belonging to exponential families (cf. [2]; [4]; [17], pp. 64, 68; [24]); more particularly, when the restrictions on the parameter point are expressed by inequalities which are linear in its coordinates.

Let $F(x)$ be a distribution function. The integral

$$\phi(\tau) = \int_{-\infty}^{\infty} e^{x\tau} \, dF(x),$$

giving its moment-generating function, converges to 1 for $\tau = 0$; we shall suppose its interval of convergence contains the origin as an interior point. It then con-

verges in a vertical strip of the complex τ plane containing the origin to an analytic function which is positive on the real axis. We set

$$\Theta(\tau) = \log \phi(\tau),$$

using the principal value of the logarithm (which is real when $\phi(\tau) > 0$, hence when τ is real); $\Theta(\tau)$ is analytic for real τ in the interval of convergence.

DEFINITION. The distribution functions $F(x; \tau)$ form an *exponential family*, the *family of exponential type determined by* $F(x)$ *or by* $\Theta(\tau)$, if, for τ in the interval of convergence,

$$F(x; \tau) = \int_{(-\infty, x)} \exp [u\tau - \Theta(\tau)] \, dF(u). \tag{4.1}$$

It was shown by Koopman [20] and by Pitman [23] that, except for change in variable or change in parameter, a (sufficiently regular) one-parameter family of distributions over a common, fixed (possibly infinite) interval admits a sufficient statistic *only* if the parameter enters as does the parameter τ in (4.1). Further, it is clear from the derivation in [11] of the Cramer-Rao inequality that in the above statement the term "sufficient" may be replaced by "efficient" (as defined in [11]).

If X_τ is a random variable whose distribution function $F(x; \tau)$ is given by (4.1), then its expectation and variance are given by

$$E(X_\tau) = \theta(\tau), \; V(X_\tau) = \theta'(\tau), \tag{4.2}$$

where

$$\theta(\tau) = \Theta'(\tau). \tag{4.3}$$

Since $V(X_\tau) \geqq 0$ it follows that $\theta(\tau)$ is increasing and $\Theta(\tau)$ convex; indeed, $\theta(\tau)$ is *strictly* increasing, and $\Theta(\tau)$ *strictly* convex unless $F(x)$ is degenerate, a possibility we shall rule out from further consideration.

We define $\tau(\theta)$ as the inverse function of $\theta(\tau)$, and $T(\theta)$ by

$$T(\theta) = \int_{\theta_0}^{\theta} \tau(v) \, dv,$$

where $\theta_0 = \theta(0)$. Evidently $T(\theta)$ is convex, and assumes its minimum value, 0, at θ_0. According to an inequality of W. H. Young ([18], p. 111), we have

$$T(x) + \Theta(y) - xy \geqq 0, \tag{4.4}$$

with equality holding if and only if $y = \tau(x)$ $(x = \theta(y))$. This becomes geometrically obvious on interpreting T and Θ relative to the graph of $y = \tau(x)$ or $x = \theta(y)$ in the xy plane.

We note that (i) a normal distribution with variable mean and fixed standard deviation, (ii) a Poisson distribution with variable mean, (iii) the distribution of the square of a normally distributed random variable having zero mean and variable variance, (iv) a binomial distribution with variable mean, and (v) a

negative binomial distribution with variable parameter p, are examples of exponential families. If X_0 is any random variable whose moment generating function exists on an open interval containing the origin, there is an exponential family of distributions admitting the distribution of X_0 as a member for one parameter value. In random sampling from a population of this family, the sample mean is the MLE of $E(X_\tau)$ (cf. discussion of (4.6) below); it is also the least squares estimator; it is unbiased, consistent, sufficient, and efficient.

Let us now consider an estimation problem. Let k be a positive integer. For $i = 1, 2, \cdots, k$, consider a population whose distribution belongs to the exponential family determined by a given distribution function $F_i(x)$ for a particular parameter value τ_i, regarded as unknown. Let $z_i = (x_{1,i}, x_{2,i}, \cdots, x_{n_i,i})$ denote the set of sample values of a sample of size n_i from the ith population, and set $z = (z_1, \cdots, z_k)$. Let $\bar{x}_i$ denote the sample mean ($i = 1, 2, \cdots, k$), and let $\bar{x}$ denote the point $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$ in the Euclidean space R_k of k dimensions. If E is an event in the sample space, its probability is given by

$$P_\tau(E) = \int_E \exp\left\{\sum_{i=1}^k n_i[\bar{x}_i \tau_i - \Theta_i(\tau_i)]\right\} dP_0(z), \tag{4.5}$$

where

$$P_0(E) = \int_E \prod_{i=1}^k \prod_{j=1}^{n_i} dF_i(x_{j,i}).$$

Set $\tau = (\tau_1, \cdots, \tau_k)$, $y = (y_1, \cdots, y_k)$. The MLE of τ is that point $y = \tau^*$ which maximizes $\sum_{i=1}^k n_i[\bar{x}_i y_i - \Theta_i(y_i)]$; or equivalently, which minimizes

$$G(y) \equiv \sum_{i=1}^k n_i[T_i(\bar{x}_i) + \Theta_i(y_i) - \bar{x}_i\, y_i]. \tag{4.6}$$

This function is convex in y. It is clear from inequality (4.4) and the remark following it that the unrestricted minimum is afforded by $\tau^* = (\tau_1^*, \tau_2^*, \cdots, \tau_k^*)$, where $\tau_i^* = \tau_i(\bar{x}_i)$ (the special case $k = 1$ was mentioned above). Suppose that restrictions on τ may be expressed by $\tau \in A_1 A_2 \cdots A_N$, where A_i is the closure of an open convex subset of $R_k (i = 1, 2, \cdots, N)$. We consider now the subproblem of minimizing G on a given intersection of boundaries of some of the sets A_i. Assuming the boundaries of the sets A_i sufficiently regular, if the unrestricted minimum of G is attained outside A_j, then the point $\tau^* = (\tau_1^*, \cdots, \tau_k^*)$ at which G assumes its minimum on $\mathcal{B}(A_j)$ satisfies

$$\sum_{\nu=1}^k \alpha_\nu^r(\tau^*) n_\nu[\theta_\nu^* - \bar{x}_\nu] = 0, r = 1, 2, \cdots, k - 1, \tag{4.7}$$

where, for $r = 1, 2, \cdots, k - 1$, $\alpha^r(\tau^*) = [\alpha_1^r(\tau^*), \cdots, \alpha_k^r(\tau^*)]$ is one of $k - 1$ independent vectors tangent at τ^* to $\mathcal{B}(A_j)$, and where $\theta_i^* = \theta_i(\tau_i^*)$. Similarly, the condition that G assume its minimum on an "edge" $\mathcal{B}(A_{i_1})\mathcal{B}(A_{i_2}) \cdots \mathcal{B}(A_{i_n})$ is (4.7) for $r = 1, 2, \cdots, k - n$, where $\alpha^r(\tau^*)$ is one of $k - n$ independent vectors tangent at τ^* to $\mathcal{B}(A_{i_1})\mathcal{B}(A_{i_2}) \cdots \mathcal{B}(A_{i_n})$. Thus the point minimizing G

on a given boundary or intersection of boundaries is a solution of equations of form (4.7). If, in particular, the boundaries $\mathcal{B}(A_i)$ are all hyperplanes, then the α_i^r are constant on a given intersection of boundaries, and values θ_i^* of the means corresponding to the coordinates τ_i^* of the minimizing point are solutions of linear equations of the form

$$\sum_{i=1}^{k} \alpha_i^r n_i[\theta_i^* - \bar{x}_i] = 0.$$

If the restricting conditions require that $\theta = (\theta_1, \cdots, \theta_k)$, rather than τ, belong to the intersection of closed convex sets, their maps in the τ-space need not in general be convex, and the above discussion need not apply. There are a number of situations of interest, however, in which the above technique for finding the minimizing point on a given intersection of boundaries will still be applicable.

(i) The function of θ, $G[\tau(\theta)]$, obtained by replacing y in (4.6) by $\tau(\theta)$, may be convex in θ. For example, this will be the case if each population is normal with known variance, or binomial, or Poisson. Since the transformation from θ-space to τ-space is 1–1 and analytic, the above discussion for finding the minimizing point in τ-space will apply even though the restricting sets in τ-space may not be convex.

(ii) All populations belong to the same exponential family, and only order restrictions are made on the parameters; that is, the regions A_i are defined by inequalities of the form $\theta_r \leqq \theta_s$. In this case $\Theta(\tau) \equiv \Theta_i(\tau)$ is independent of i, and $\tau_r \leqq \tau_s$ if and only if $\theta_r = \theta(\tau_r) \leqq \theta(\tau_s) = \theta_s$; since $\theta(\tau)$ and $\tau(\theta)$ are strictly increasing. The independent vectors α^r for a given "edge" in this case are determined by the indices i of the boundaries intersecting in the edge, independently of the particular function. *The* MLE (cf. Section 6 for a specific description in a special case) *of θ is therefore independent of the particular exponential family to which the populations belong, provided they all belong to the same exponential family, and provided only order restrictions are made on the parameters* θ_i, $i = 1, 2, \cdots, k$. In particular, for the purpose of determining the MLE's of the means, one could in such a situation assume without loss of generality that the populations are all normal with standard deviation 1, but with possibly different means, satisfying the specified order restrictions. (In the special case where the order restrictions specify a simple ordering of the means, the failure of the MLE's to depend on the particular exponential family was noted in [6] and in [7]). Thus in this situation the problem of finding the MLE reduces to that of minimizing the function

$$\sum_{i=1}^{k} n_i(\bar{x}_i - \theta_i)^2$$

subject to specified restrictions of the form $\theta_r \leqq \theta_s$. With an obvious linear change of variable, it can be expressed as the problem of finding the foot of the segment of smallest length from a given point onto a set bounded by hyperplanes passing through the origin.

5. A sufficiency property. Let us consider for a moment the simplest case of estimating a restricted parameter. We sample from a single population, belonging to an exponential family. The parameter θ is known to lie in a proper subinterval of its natural range. The MLE, $\bar{x}$, of the unrestricted parameter is known to be consistent, efficient, sufficient, and unbiased. It seems to the author that a "reasonable" estimator of the restricted parameter is $\bar{x}$, appropriately truncated, which is also the MLE. This estimator is not sufficient (nor unbiased). Likewise, in the more general situation discussed in Section 4, the MLE is not sufficient. However, it does possess a certain "sufficiency-like" property, expressed in Theorem 5.1. Referring to the general problem formulated in Section 4, we suppose that the parameter point $\tau = (\tau_1, \cdots, \tau_k)$ is subject to the restriction $\tau \varepsilon S_0 = A_1A_2, \cdots, A_N$, where now each A_j is a closed set bounded by a hyperplane. (In the event that all populations are normal with the same standard deviation, or that all populations belong to the same exponential family and the equation of the boundary of each A_j is of the form $\tau_r \leqq \tau_s$, the corresponding sets in θ-space will also be bounded by hyperplanes.) Let z denote a point of the sample space, and let $Y(z) = [Y_1(z), Y_2(z), \cdots, Y_k(z)]$ denote the corresponding MLE of τ, subject to $\tau \varepsilon S_0$. For a Borel set E in the sample space, let $p_\tau(E \mid y)$ denote the conditional probability of E for a given value y (in S_0) of $Y(z)$. That is, $p_\tau(E \mid y)$ is to be defined so that for each Borel set $B \subset S_0$ we have

$$P_\tau(E \cap Y^{-1}(B)) = \int_B p_\tau(E \mid y)\, dP_\tau Y^{-1}(y), \tag{5.1}$$

where $P_\tau(E)$ is given by (4.5) for each event E in the sample space, where $Y^{-1}(B)$ denotes the inverse image of B under the map Y from the sample space into S_0, and where $P_\tau Y^{-1}(B) = P_\tau[Y^{-1}(B)]$

THEOREM 5.1. *Let S_0 be bounded by hyperplanes. There is a determination of $p_\tau(E \mid y)$ which is independent of τ when y is interior to S_0, and, when y lies interior to a $(k-1)$-dimensional face or $(k-j)$-dimensional ($j = 2, 3, \cdots, k$) edge or vertex of S_0, is independent of τ on the closure of that face, edge, or vertex.*

PROOF. For x in θ-space, define $y(x)$ by $y(x) = (\tau_1(x_1), \tau_2(x_2), \cdots, \tau_k(x_k))$. For z in the sample space, define $V(z) = y(\bar{x})$. We have $Y(z) = V(z)$ if $y(\bar{x}) \varepsilon S_0$. Define $q(E \mid y)$ to be the conditional probability of E given a value y of $V(z)$; this conditional probability may be taken to be independent of τ, since $V(z)$ is a sufficient estimator of τ. For y interior to S_0, we define

$$p_\tau(E \mid y) = q(E \mid y), \qquad \text{for all } \tau \varepsilon S_0.$$

Then if B is interior to S_0, and if $\tau \varepsilon S_0$, we have

$$P_\tau(E \cap Y^{-1}(B)) = P_\tau(E \cap V^{-1}(B)) = \int_B q(E \mid y)\, dP_\tau V^{-1}(y)$$

$$= \int_B p_\tau(E \mid y)\, dP_\tau Y^{-1}(y).$$

Now suppose y is on a $(k - 1)$-dimensional face or $(k - j)$-dimensional ($j = 2, 3, \cdots, k$) edge, W, of S_0, which is open in its relative topology. For τ not on the closure, W^{cl}, of W, let $p_\tau(E \mid y)$ denote any determination of the conditional probability satisfying (5.1). Choose a fixed $\beta \varepsilon W$, and let $p_\beta(E \mid y)$ denote any determination of the conditional probability satisfying (5.1). For τ on W^{cl}, define $p_\tau(E \mid y)$ to be equal to $p_\beta(E \mid y)$. We now wish to verify that p so defined satisfies (5.1) when B is a Borel subset of W. For such B we have, by definition,

$$P_\beta[E \cap Y^{-1}(B)] = \int_{E\cap Y^{-1}(B)} \exp\left\{\sum_{i=1}^{k} n_i[\bar{x}_i \beta_i - \Theta_i(\beta_i)]\right\} dP_0(z)$$

$$= \int_B p_\beta(E \mid y)\, dP_\beta Y^{-1}(y).$$

Also

$$P_\tau[E \cap Y^{-1}(B)] = \int_{E\cap Y^{-1}(B)} \exp\left\{\sum_{i=1}^{k} n_i \bar{x}_i(\tau_i - \beta_i) - n_i[\Theta_i(\tau_i) - \Theta_i(\beta_i)]\right\}\cdot$$

$$\exp\left\{\sum_{i=1}^{k} n_i[\bar{x}_i \beta_i - \Theta_i(\beta_i)]\right\} dP_0(z).$$

If the MLE, Y, of τ is in W, then, by (4.7),

$$\sum_{i=1}^{k} \alpha_i^r n_i \bar{x}_i = \sum_{i=1}^{k} \alpha_i^r n_i \theta_i(Y_i),$$

where the α^r are independent vectors spanning W. If $\tau \varepsilon W^{cl}$, then $\tau - \beta$ is a linear combination of the α^r; hence

$$\sum_{i=1}^{k} n_i \bar{x}_i(\tau_i - \beta_i) = \sum_{i=1}^{k} n_i \theta_i(Y_i)(\tau_i - \beta_i),$$

a function of Y for fixed τ, β. So also, then, is

$$\exp\left\{\sum_{i=1}^{k} n_i \bar{x}_i(\tau_i - \beta_i) - [\Theta_i(\tau_i) - \Theta_i(\beta_i)]\right\}$$

a function, $\psi[Y(z)]$, of $Y(z)$. We have then

$$P_\tau[E \cap Y^{-1}(B)] = \int_{E\cap Y^{-1}(B)} \psi[Y(z)]\, dP_\beta(z)$$

$$= \int_B \psi(y) p_\beta(E \mid y)\, dP_\beta Y^{-1}(y)$$

$$= \int_B p_\beta(E \mid y)\, dP_\tau Y^{-1}(y),$$

since

$$dP_\tau Y^{-1}(y) = \psi(y)\, dP_\beta Y^{-1}(y).$$

One now verifies from the appropriate definitions above that (5.1) holds for arbitrary $\tau \varepsilon S_0$, and Borel set $B \subset S_0$. This completes the proof of Theorem 5.1.

Theorem 5.1 *may be regarded as a generalization of a remark* ([16], p. 77) to the effect that if X and Y are normally distributed random variables with unit standard deviation and means ξ and η respectively, and if ξ and η are known to satisfy a linear equation, then the foot of the perpendicular from the observation *point* (x, y) *is a sufficient estimator.*

Theorem 5.1 may be interpreted somewhat as follows. Given the value of $Y(z)$, the exact knowledge of the observed sample point would imply no additional information as to how to select τ on the face (or edge) on which $Y(z)$ lies, *since the conditional distribution, given* $Y(z)$, *is independent of* τ *on this* face.

6. Uniform consistency of a class of estimators. In Section 4, an estimation problem of the following kind was considered. Let k be a positive integer. To each positive integer $i \leqq k$ corresponds a population whose distribution is known except for the unknown value of its mean, θ_i. The means θ_i are known to satisfy certain inequalities. The problem of estimating a distribution function from all-or-none data (bioassay) is of this kind, in which the populations are binomial and the inequalities are of the form $\theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_k$ (cf. Section 3; also [1], [12]). Even if the populations are not binomial, but all belong to a common exponential family, the MLE's subject to $\theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_k$ are very easily determined, as follows (cf [1], [7]). Let $\bar{x}_i$ denote the sample mean of a sample of size n_i from the i-th population, whose mean is θ_i, $i = 1, 2, \cdots, k$. If $\bar{x}_1 \leqq \bar{x}_2 \leqq \cdots \leqq \bar{x}_k$, these are the MLE's of the parameters θ_i, $i = 1, 2, \cdots, k$ If for some i we have $\bar{x}_i > \bar{x}_{i+1}$, these two means are replaced by the single ratio $(n_i\bar{x}_i + n_{i+1}\bar{x}_{i+1}) / (n_i + n_{i+1})$, obtaining an ordered set of only $k - 1$ ratios ($k - 2$ of which are sample means). This procedure *is* repeated *until* an ordered set of ratios is obtained which are monotone non-decreasing Then for each i, the MLE, $\hat{\theta}_i$, of θ_i is equal to that one of the final set of ratios to which the original ratio $\bar{x}_i$ contributed.

If the number, k, of *observation points is held fixed*, while *the number of observations* at each point increases indefinitely, classical theory assures the strong consistency of the $\hat{\theta}$, and yields their asymptotic distribution; the $\hat{\theta}$ will asymptotically coincide with the sample means. We shall be interested here chiefly in *situations in which there are a large number of observation points, but only a* few observations, perhaps only one, at each. In [1] and in [7] the *local consistency* of the MLE's is proved. It is assumed that there is an unknown function $\theta(t)$ (as in bioassay, for example), known to be non-decreasing and continuous, *such that* $\theta_i = \theta(t_i)$, $i = 1, 2, \cdots, k$. *Then if t is held fixed, one can achieve an* arbitrarily high probability of an arbitrarily great precision at t by selecting enough observation points in the neighborhood of t, even if only one observation is made at each. In [1] and [7] it was assumed that the populations all belonged *to the same exponential family; but it is clear that the estimators $\hat{\theta}$ can be formed*

without regard to the distributions of the k populations; they are determi by the sample means alone (of course, they will not in general be MLE Indeed, the proof of the local consistency of the estimators $\hat{\theta}$ does not req an assumption that the populations belong to an exponential family.

Theorem 6.2 below gives conditions sufficient for the *strong uniform consist* of the estimators $\hat{\theta}$, *without assuming the populations belong to an exponen family*. The proof requires a somewhat strengthened form of the strong lav large numbers, which is presented in Theorem 6.1.

THEOREM 6.1. *Let r be a fixed positive number. Let $Y_1, Y_2, \cdots,$ be indepen random variables with $E(Y_i) = 0$, $E(|Y_i|^{2r}) < \infty$, and*

$$\sum_i E(|Y_i|^{2r})/i^{r+1} < \infty. \tag{6.1}$$

Corresponding to each positive integer $n \geqq 2$, let $i_{1,n}, i_{2,n}, \cdots, i_{n,n}$ be a perm tion of the positive integers $1, 2, \cdots, n$, obtained by assigning a place to the in n between some two successive integers, or at the beginning, or at the end, o permutation corresponding to the integer $n - 1$. Define $S_{j,n} = \sum_{\nu=1}^{j} Y_i$ $j = 1, 2, \cdots, n$. Then

$$\Pr\left\{\lim_{n\to\infty} \max_{j=1,2,\cdots,n} \frac{1}{n} |S_{j,n}| = 0\right\} = 1.$$

INDICATION OF PROOF OF THEOREM 6.1. The situation is more complic than that of the classical strong law, but familiar arguments suffice. For ν 1, $\cdots$, arrange the terms Y_i having indices i such that $2^{\nu-1} < i < 2^{\nu}$ in order given by the permutation for 2^{ν}, and let $\mathfrak{J}(\nu)$ denote the family of pa sums containing the first of these terms, the sum of the first two, the sum o first three, etc. Now consider partial sums $S_{j,n}$, $j \leqq n$. For each n, choose $k =$ so that $2^{k-1} < n \leqq 2^k$. To avoid complicated subscripts, let $p = p(n) =$ Let $Z_1, Z_2, \cdots, Z_{2p-n}$ denote the random variables $Y_{n+1}, Y_{n+2}, \cdots,$ written in the order given by the permutation for $2p = 2^k$. Let $\mathfrak{U}(n)$ de the family of partial sums: $\{Z_1, Z_1 + Z_2, \cdots, Z_1 + Z_2 + \cdots + Z_{2p-n}\}$ fixed j, n, and for $\nu = 0, 1, 2, \cdots, k - 1$, let $T_\nu = T_\nu(j, n)$ denote the su terms Y_i which appear in the sum $S_{j,n}$ and which have indices i such $2^{\nu-1} < i \leqq 2^{\nu}$. Then $T_\nu \,\varepsilon\, \mathfrak{J}(\nu)$ for $\nu = 0, 1, 2, \cdots, k - 1$. Let $T_k = T_k(j, n)$ de the minimal member of $\mathfrak{J}(k)$ containing all terms appearing in $S_{j,n}$ whose ind satisfy $2^{k-1} < i \leqq 2^k$ (minimal in the sense of containing the fewest pos terms). Let $U = U(j, n)$ be the sum of terms appearing in T_k of index gr than n; then $U \,\varepsilon\, \mathfrak{U}(n)$, and $S_{j,n} = \sum_{\nu=0}^{k} T_\nu - U$. Let $\mathfrak{V}(k)$ denote the fa of all sums of the form $\sum_{\nu=0}^{k} W_\nu$, where $W_\nu \,\varepsilon\, \mathfrak{J}(\nu)$, $\nu = 0, 1, 2, \cdots, k$. $V = V(j, n) = \sum_{\nu=0}^{k} T_\nu$. Then $V \,\varepsilon\, \mathfrak{V}(k)$ and

$$S_{j,n} = V - U.$$

Let ϵ be positive. Let A_n denote the event: $\{\max_{0\leqq j\leqq n}|S_{j,n}| > 2^{k+1}\epsilon\}$, B event: $\{\max_{0\leqq j\leqq n} |U| \leqq 2^k\epsilon\}$, and C_k the event: $\{\max_{V\varepsilon\mathfrak{V}(k)} |V| > 2^k\epsilon\}$.

$$A_nB_n \subset C_k \qquad (k = k \tag{6.2}$$

It follows from Chung's inequality ([10], p. 348) and the generalized Kolmogorov inequality ([21], p. 265), that

$$P(B_n) > 1 - A\left[\sum_{\nu=p+1}^{2p} E(|Y_\nu|^{2r})\right] \Big/ (2^k\epsilon)^{r+1},$$

where A is a constant depending only on r. From hypothesis (6.1), using Kronecker's lemma, we conclude that

$$\lim_{k\to\infty}\left[\sum_{\nu=p+1}^{\infty} E(|Y_\nu|^{2r})\right] \Big/ (2^k\epsilon)^{r+1} = 0,$$

hence there is a positive integer k_0 such that $P(B_n) > \frac{1}{2}$ for $n > p_0 = 2^{k_0-1}$. Further, A_{p_0+1} and B_{p_0+1} are independent, and if A^c denotes the complement of A, then for $n = p_0 + 2, p_0 + 3, \cdots$, we have that $A^c_{p_0+1} \cap A^c_{p_0+2} \cap \cdots A^c_{n-1} \cap A_n$ and B_n are independent. It follows from the "Lemma for Events", [21], p 246, that

$$P(\cup_{n=p_0+1}^{\infty} A_n B_n) \geqq \tfrac{1}{2} P(\cup_{n=p_0+1}^{\infty} A_n),$$

so that from (6.2) we have

$$P(\cup_{n=p_0+1}^{\infty} A_n) \leqq 2P(\cup_{\nu=k_0}^{n} C_\nu),$$

hence

$$P(\limsup_{n\to\infty} A_n) \leqq 2\, P(\limsup_{\nu\to\infty} C_\nu),$$

or

$$\Pr\{\max_{0\leq j\leq n} |S_{j,n}| > 2^{k+1}\epsilon \quad \text{for infinitely many } n\}$$

$$\leqq 2 \Pr\{\max_{V\epsilon\mathfrak{V}(\nu)} |V| > 2^\nu\epsilon \quad \text{for infinitely many } \nu\}.$$

Kolmogorov's method ([19], cf. also [25], p. 202), with Chung's inequality and the generalized Kolmogorov inequality can be used to show that the right hand member is 0. Since $2^{k+1} < 4n (k = k(n), 2^{k-1} < n \leqq 2^k)$, we have

$$\Pr\left\{\max_{0\leq j\leq n} \frac{1}{n}\left|S_{j,n}\right| > 4\epsilon \quad \text{for infinitely many } n\right\} = 0$$

A standard argument completes the proof.

We return now to the estimation problem. For $i = 1, 2, \cdots, k$, $\bar{x}_i$ is the sample mean of a sample of size n_i from a population whose mean is $\theta(t_i)$. It is known that $\theta(t)$ is non-decreasing. We are concerned with the estimator $\hat{\theta}(t)$ obtained as described above. It is given ([1], [7]) by

$$(6.3)\qquad \begin{cases} \hat{\theta}(t) = \max\limits_{t_r\leq t}\min\limits_{t_s\geq t}\left(\sum\limits_{\nu=r}^{s} n_\nu \bar{x}_\nu\right) \Big/ \left(\sum\limits_{\nu=r}^{s} n_\nu\right) \\ \quad\;\; = \min\limits_{t_s\geq t}\max\limits_{t_r\leq t}\left(\sum\limits_{\nu=r}^{s} n_\nu \bar{x}_\nu\right) \Big/ \left(\sum\limits_{\nu=r}^{s} n_\nu\right). \end{cases}$$

THEOREM 6.2. *Let $\theta(t)$ be continuous and non-decreasing on (a, b). Let $\{s_n\}$ be a sequence of observation points dense in (a, b). Let one observation be made at each point (the observation points need not be distinct). Let the variances of the observed random variables be bounded. Let $\hat{\theta}_n(t)$ denote the estimate of $\theta(t)$ based on observations made at the first n observation points, defined to be constant between observation points, and continuous from the left. If $c > a$, $d < b$, then*

$$\Pr\{\lim_{n\to\infty} \max_{c\leq t\leq d} |\hat{\theta}_n(t) - \theta(t)| = 0\} = 1.$$

PROOF. The original proof used Theorem 6.1 and a geometrical interpretation of $\hat{\theta}$ due to W. T. Reid [5] which is also used in the proof of Theorem 6.3. It required as additional hypothesis that the norm (maximum distance between adjacent points of subdivision) of the subdivision of (a, b) formed by the first n observation points be $O(1/n)$, and required the less restrictive hypothesis (6.1) on the variances of the observable random variables. The present proof uses an approach suggested by the referee. This proof also could be modified to use the hypothesis (6.1) on the variances instead of boundedness, together with a uniformity condition on the distribution of the observation points, but the above formulation appears more natural and useful.

We observe first that if $\hat{\theta}_n(u_i) - \theta(u_i) \to 0$ for each u_i of a sequence $\{u_i\}$ dense in (a, b), then it follows from the monotonicity of $\hat{\theta}_n$ and the continuity of θ that $\max_{c\leq t\leq d} |\hat{\theta}_n(t) - \theta(t)| \to 0$. Consequently it suffices to show that, for each individual $t \in (a, b)$, $\Pr\{\hat{\theta}_n(t) - \theta(t) \to 0\} = 1$, since it then follows that $\Pr\{\hat{\theta}_n(u_i) - \theta(u_i) \to 0 \text{ for all } u_i\} = 1$, if $\{u_i\}$ is any countable sequence of points in (a, b).

We now prove that for fixed $t \in (a, b)$, $\Pr\{\hat{\theta}_n(t) - \theta(t) \to 0\} = 1$. It suffices to prove that for every $\epsilon > 0$ we have

$$\Pr\{\liminf_{n\to\infty} \hat{\theta}_n(t) - \theta(t) \geqq -\epsilon\} = 1 \tag{6.4}$$

and

$$\Pr\{\limsup_{n\to\infty} \hat{\theta}_n(t) - \theta(t) \leqq \epsilon\} = 1. \tag{6.5}$$

We prove the first; the proof of the second is similar.

We suppose the sequence $\{s_i\}$ of observation points chosen, not necessarily distinct nor ordered according to increasing index, and an observation Z_i made at each, so that $E(Z_i) = \theta(s_i)$. Let $\sigma_i^2 = V(Z_i)$, the variance of the random variable Z_i observable at s_i. For fixed n, let $t_1, t_2, \cdots, t_k$ denote the $k = k(n)$ distinct observation points among $s_1, s_2, \cdots, s_n$, arranged in increasing order, and let n_i denote the number of observations made at t_i, so that $\sum_{i=1}^k n_i = n$.

Let $t \in (a, b)$. Given $\epsilon > 0$, choose n sufficiently large that there is a $t_r < t$ such that $|\theta(t_r) - \theta(t)| < \epsilon$. By (6.3),

$$\hat{\theta}_n(t) - \theta(t) \geqq \min_{t_s\geqq t} \frac{\sum_{\nu=r}^{s} n_\nu[\bar{x}_\nu - \theta(t_r)]}{\sum_{\nu=r}^{s} n_\nu} - [\theta(t) - \theta(t_r)].$$

Since θ is non-decreasing, we have $\theta(t_r) \leqq \theta(t_\nu)$ for $\nu \geqq r$, hence

(6.6) $$\hat{\theta}_n(t) - \theta(t) > \min_{t_s \geq t} \frac{\sum_{\nu=r}^{s} n_\nu[\bar{x}_\nu - \theta(t_\nu)]}{\sum_{\nu=r}^{s} n_\nu} - \epsilon.$$

For $\rho = 1, 2, \cdots$, let s_{i_ρ} denote the ρth of the members of the sequence $\{s_i\}$ which lie at, or to the right of, t_r. Consider the sequence of observable random variables, centered at means, $Z_{i_\rho} - \theta(s_{i_\rho})$. The sums $\sum_{\nu=r}^{s} n_\nu[\bar{x}_\nu - \theta(t_\nu)]$ are *not* successive partial sums of this sequence or of any sequence, since as ρ increases new observation points are interspersed among the old. However, in applying Theorem 6.1 with $Y_\rho = Z_{i_\rho} - \theta(s_{i_\rho})$, we find that the ratios $\sum_{\nu=r}^{s} n_\nu[\bar{x}_\nu - \theta(t_\nu)] / \sum_{\nu=r}^{s} n_\nu$ are just such ratios $S_{\nu,n} / n$ as are considered there. We conclude from (6.6) that $\Pr\{\liminf_{n\to\infty} \hat{\theta}_n(t) - \theta(t) \geqq -\epsilon\} = 1$. A similar argument shows that for $\epsilon > 0$,

$$\Pr\{\limsup_{n\to\infty} \hat{\theta}_n(t) - \theta(t) \leqq \epsilon\} = 1,$$

whence

$$\Pr\{\lim_{n\to\infty} \hat{\theta}_n(t) - \theta(t) = 0\} = 1.$$

Together with the earlier remarks, this completes the proof of Theorem 6.2.

Theorem 6.3, below, gives an asymptotic lower bound for the probability of achieving a given uniform precision on a closed subinterval of (a, b).

THEOREM 6.3. *For a fixed positive integer n, let n observations be made at observation points $t_1 \leqq t_2 \leqq \cdots \leqq t_k$ in (a, b), n_i observations being made at t_i, $i = 1, 2, \cdots, k$, so that $n = \sum_{i=1}^{k} n_i$. Let $\Delta = \max_{i=0,1,2,\cdots,k}(t_{i+1} - t_i)$, $t_0 = a$, $t_{k+1} = b$. Let the populations be such as to permit the application of the Central Limit Theorem as required in* [15] (cf. *also* [9]; *for an appropriate Lindeberg condition, see* [22], p. 127). *Let $\theta(t)$ have a bounded derivative, $|\theta'(t)| \leqq K$, $K > 0$, for $t \in (a, b)$, and let $\sigma^2 = \sum_{i=1}^{k} n_i\sigma_i^2$, where σ_i^2 is the variance of an observation made at t_i, $i = 1, 2, \cdots, k$. For $z > 0$, let $h = [2z\sigma\Delta / K]^{\frac{1}{2}}$, $c = a + h$, $d = b - h$. Then*

$$\Pr\{\max_{c\leq t\leq d} |\hat{\theta}(t) - \theta(t)| < 2\sqrt{2Kz\sigma\Delta}\} \gtrsim \frac{4}{\pi}\sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{2\nu+1} \exp[-(2\nu+1)^2\pi^2/8z^2].$$

The symbol "$\gtrsim$" is to be interpreted as "asymptotically (as $n \to \infty$) at least as large as". The estimate is most nearly accurate if only one observation is made at each point, if the observation points are distributed uniformly over (a, b), and if $\theta'(t)$ is constant.

PROOF. Order the observations according to increasing t, ordering in an arbitrary way those occurring at the same observation point. Let $Z_{\nu,n}$ denote the νth observation, $\nu = 1, 2, \cdots, n$; its mean is $\theta(t_j)$ and its variance σ_j^2 if it is made at the observation point t_j. For positive integers $j \leqq k$, define $N_j = \sum_{t_\nu \leq t_j} n_\nu$, $s(N_j) = \sum_{t_\nu \leq t_j} n_\nu\theta(t_\nu)$, and $s^*(N_j) = \sum_{t_\nu \leq t_j} n_\nu\bar{x}_\nu$. We have $s^*(N_j)$ as one of the partial sums of the sequence $Z_{\nu,n}$, and $s(N_j)$ as its expecta-

tion. If $S_{\nu,n}$ denotes the νth partial sum, and s_ν^2 its variance, then it is known that

$$\lim_{n\to\infty} \Pr\{\max_{\nu\leqq n} |S_{\nu,n} - E(S_{\nu,n})| < zs_n\} = \frac{4}{\pi}\sum_{i=0}^{\infty} \frac{(-1)^i}{2i+1} \exp[-(2i+1)^2\pi^2/8z^2]$$

(strictly speaking, we require that the theorem as developed in [15] and [9] be generalized so as to apply to sums of the form $\sum_{\nu=1}^{n} X_{\nu,n}$, where the $X_{\nu,n}$ are independent for distinct ν, rather than to sums of the form $\sum_{\nu=1}^{n} X_\nu$; but only trivial modifications are required in the proofs). Define $s(u)$ and $s^*(u)$ to be linear between successive integers; then

$$(6.7)\quad \Pr\{\max_{0\leqq u\leqq n} |s^*(u) - s(u)| \leqq z\sigma\} \mathrel{\dot{\geqq}} \frac{4}{\pi}\sum_{i=0}^{\infty} \frac{(-1)^i}{2i+1} \exp[-(2i+1)^2\pi^2/8z^2].$$

We observe that $s(u)$ is a convex function whose graph consists of line segments: for $N_{i-1} < u < N_i$ we have $s'(u) = \theta(t_i)$, $i = 1, 2, \cdots, k$. The graph of $s^*(u)$ also consists of line segments, but it need not be convex, since $\bar{x}_i$ need not increase with i.

Let $g(u)$ denote the greatest convex function not greater than $s^*(u)$; the graph of this function consists of line segments. We denote by $g'(u)$ ($s'(u)$) the derivative of $g(u)$ ($s(u)$) where it is defined, and the left-hand limit of the derivative at a corner. One verifies from formulas (6.3) that $\hat{\theta}(t_i) = g'(N_i)$, $i = 1, 2, \cdots, k$. Now let u' be fixed, so that $u' \leqq \sum_{t_\nu\leqq d} n_\nu$. Suppose $\max_{0\leqq u\leqq n} |s^*(u) - s(u)| < z\sigma$. Then for $u \geqq u'$ we have

$$g(u) \leqq s^*(u) < s(u) + z\sigma.$$

Since the point $(u', g(u'))$ is on a line segment whose endpoints are at vertical distance less than $z\sigma$ from the graph of s (or else it is itself such an endpoint), and since s is convex, we have also

$$g(u') > s(u') - z\sigma.$$

Hence $g(u) - g(u') < s(u) - s(u') + 2z\sigma$. Therefore

$$g'(u') \leqq \frac{g(u) - g(u')}{u - u'} < \frac{s(u) - s(u')}{u - u'} + \frac{2z\sigma}{u - u'}.$$

Choose i, j so that $N_{i-1} < u' \leqq N_i$, $N_{j-1} < u' \leqq N_j$. We have $[s(u) - s(u')] / (u - u') \leqq s'(u) = \theta(t_j) \leqq \theta(t_i) + K(t_j - t_i) = s'(u') + K(t_j - t_i)$. But $t_j - t_i \leqq (N_{j-1} - N_i)\,\Delta + \Delta \leqq (u - u' + 1)\Delta$, so that

$$g'(u') < s'(u') + K(u - u' + 1)\Delta + 2z\sigma / (u - u')$$

for $u' < u \leqq n$. We choose $u = u' + [2z\sigma / K\Delta]^{\frac{1}{2}}$, and find that $g'(u') - s'(u') < 2[2Kz\sigma\Delta]^{\frac{1}{2}} + K\Delta \doteq 2[2Kz\sigma\Delta]^{\frac{1}{2}}$. Similarly, $g'(u') - s'(u') \mathrel{\dot{>}} -2[2Kz\sigma\Delta]^{\frac{1}{2}}$, if $u' \geqq \sum_{t_\nu\leqq c} n_\nu$. Since $\hat{\theta}(t_i) = g'(N_i)$ and $\theta(t_i) = s'(N_i)$, $i = 1, 2, \cdots, k$, we have

$$\max_{c\leqq t\leqq d} |\hat{\theta}(t) - \theta(t)| \mathrel{\dot{<}} 2[2Kz\sigma\Delta]^{\frac{1}{2}}$$

if

$$\max_{0 \leq u \leq n} |s^*(u) - s(u)| < z\sigma.$$

The conclusion of the theorem follows from (6.7).

If $K = 0$, or if we wish a lower bound on the probability for uniform precision over a larger subinterval $[c, d]$, we must simply take u in the above discussion equal to $u' + h / \Delta$, where $h = \max [b - d, c - a]$, obtaining

$$|g'(u') - s'(u')| \leq K(h + \Delta) + 2z\sigma\Delta / h,$$

or

$$\max_{c \leq t \leq d} |\hat{\theta}(t) - \theta(t)| \leq K(h + \Delta) + 2z\sigma\Delta/h$$

To get an idea of the rate of convergence guaranteed with at least a certain probability, suppose $\theta(t) \equiv t$ on $(0, 1)$, and that $\Delta = 1 / (n + 1)$, one binomial observation being made at each observation point $i / (n + 1)$, $i = 1, 2, \cdots, n$. We find $\sigma^2 \doteq n / 6$, $K = 1$, $h = (2z^2 / 3n)^{\frac{1}{4}}$, and

$$\Pr\{\max_{c \leq t \leq d} |\hat{\theta}(t) - \theta(t)| < 2(2z^2/3n)^{\frac{1}{4}}\} \gtrapprox \frac{4}{\pi} \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{2\nu + 1} \exp\left[-\frac{(2\nu + 1)^2\pi^2}{8z^2}\right],$$

which suggests that the minimum precision (reciprocal of error) assured with a given probability increases like $n^{\frac{1}{4}}$. On the other hand, if the observations are concentrated near a given point, Theorem 3.1 of [1] suggests that the precision at that point increases like $n^{\frac{1}{3}}$.

REFERENCES

[1] Miriam Ayer, H. D. Brunk, G. M Ewing, W. T Reid, and Edward Silverman, "An empirical distribution function for sampling with incomplete information", *Ann. Math. Stat.*, Vol 26 (1955), pp. 641–647

[2] D. E. Barton, "A class of distributions for which the maximum likelihood estimate is unbiased and of minimum variance for all sample sizes", *Biometrika*, Vol 43(1956), pp. 200-202

[3] E M. L Beale, "On minimizing a convex function subject to linear inequalities", *J. Roy. Stat. Soc.*, Ser B, Vol 17(1955), pp. 173–184

[4] David Blackwell and M A Girshick "Theory of games and statistical decisions," *John Wiley and Sons, Inc*, 1954

[5] H. D Brunk, G. M. Ewing, and W T Reid, "The minimum of a certain definite integral suggested by the maximum likelihood estimate of a distribution function", (Abstract) *Bull Amer. Math Soc*, Vol. 60 (1954), p. 535

[6] H. D. Brunk, G. M. Ewing, and W. R. Utz, "Minimizing integrals in certain classes of monotone functions", *Pacific J. Math.*, Vol 7(1957), pp. 833–847

[7] H. D. Brunk, "Maximum likelihood estimates of monotone parameters", *Ann Math Stat.*, Vol. 26(1955), pp. 607–616.

[8] H. D Brunk, "The strong law of large numbers", *Duke Math J*, Vol. 15(1948), pp 181–195.

[9] K. L. Chung, "On the maximum partial sums of sequences of independent random variables", *Trans. Amer. Math Soc.*, Vol. 64 (1948), pp 205–233

[10] K. L. CHUNG, "The strong law of large numbers", *Proc. Second Berkeley Symposium on Math. Stat. and Prob.*, Univ. of California Press, 1951, pp. 341–352.
[11] HARALD CRAMÉR, "Mathematical Methods of Statistics, *Princeton University Press*, 1946.
[12] W. F. DIXON AND F. J. MASSEY, "Introduction to Statistical Analysis", *McGraw-Hill*, 1951.
[13] CONSTANCE VAN EEDEN, "Maximum likelihood estimation of ordered probabilities", *Proc. Koninkl. Nederl. Akad. van Wetenschappen, Amsterdam*, Ser. A, Vol. 59, and *Indag. Math.*, Vol. 18 (1956), pp. 444–455.
[14] CONSTANCE VAN EEDEN, "Maximum likelihood estimation of partially or completely ordered parameters", *Proc. Koninkl. Nederl. Akad. van Wetenschappen, Amsterdam*, Ser. A, Vol. 60 and *Indag. Math.*, Vol. 19 (1957), pp. 128–211.
[15] P. ERDÖS AND M. KAC, "On certain limit theorems of the theory of probability", *Bull. Amer. Math. Soc.*, Vol. 52 (1946), pp. 292–302.
[16] R. A. FISHER, "Statistical methods and scientific induction", *J. Roy. Stat. Soc.*, Ser. B, Vol. 17 (1955), pp. 69–78.
[17] M. A. GIRSHICK AND L. J. SAVAGE, "Bayes and minimax estimates for quadratic loss functions", *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 53–73.
[18] G. H. HARDY, J. E. LITTLEWOOD, AND G. PÓLYA. "Inequalities," *Cambridge University Press*, Second edition, 1952.
[19] A. KOLMOGOROV, "Sur la loi forte des grands nombres", *Comptes Rendus Paris*, Vol. 191 (1930), pp. 910–911.
[20] B. O. KOOPMAN, "On distributions admitting a sufficient statistic", *Trans. Amer. Math. Soc.*, Vol. 39 (1936), pp. 399–409.
[21] M. LOÈVE, "Probability Theory", *D. van Nostrand Co.*, 1955.
[22] W. G. MADOW, "Limiting distributions of quadratic and bilinear forms", *Ann. Math. Stat.*, Vol. 11 (1940), pp. 125–146.
[23] E. J. G. PITMAN, "Sufficient statistics and intrinsic accuracy", *Proc. Camb. Phil. Soc.*, Vol. 32 (1936), pp. 567–579.
[24] M. C. K. TWEEDIE, "Functions of a statistical variate with given means with special reference to Laplacian distributions", *Proc. Camb. Phil. Soc.*, Vol. 43 (1947), pp. 41–49.
[25] J. V. USPENSKY, "Introduction to Mathematical Probability", *McGraw-Hill*, New York, 1937.

INTERSECTION REGION CONFIDENCE PROCEDURES WITH AN APPLICATION TO THE LOCATION OF THE MAXIMUM IN QUADRATIC REGRESSION

By David L. Wallace

University of Chicago

1. Summary. Confidence region procedures for multidimensional quantities sometimes require prohibitive amounts of computation and the regions are difficult to represent in a useful way. Some approximate procedures are constructed by using regions obtained as the intersection of several regions, each much easier to construct. The procedures are applicable to the solution of simultaneous equations, whose coefficients are subject to random error. Approximations by convex polyhedra and by parallelepipeds are proposed. The procedures are illustrated for setting a confidence region for the location of the vertex of a quadratic regression surface.

2. Confidence regions. In this section, I give a subjective evaluation of the requirements for a useful confidence region procedure.

Suppose that λ is a (multidimensional) quantity defined as a function of the parameters of the distribution sampled. The problem of constructing confidence regions for the true value(s) of λ will be considered.

A confidence region and a point estimate for λ are often used to summarize the information about λ in the observed sample Their use is an attempt to convey in a comprehensible way some idea of the extent and character of the determination of λ, taking account of the inaccuracies of measurement. Any use of the confidence region in making decisions about further experimentation, process operations, etc., will be informal. The exact confidence level is not important and even the frequency interpretation of the procedure is not essential, both serving principally as "benchmarks" for purposes of comparison and familiarity. What is important is that the region be represented geometrically or analytically so that the user can comprehend its size, shape and location Approximations to the region which simplify this representation will be valuable as long as they do not greatly change the confidence level.

The theoretical specification of confidence procedures and the investigation of their statistical properties (level, power, etc.) are usually accomplished through an associated family of tests of hypotheses. The condition that the true quantity have the particular value λ is a condition on the parameters and hence a statistical hypothesis. Denote it by H_λ. Then, given a level α test of H_λ for each value of λ, the confidence procedure defined by $R = \{\lambda: H_\lambda \text{ not rejected}\}$ is an error level α confidence procedure. The "error level" (= 1 − "confidence level") of a confidence procedure is usually more convenient than the confidence level

Received November 5, 1956.

and coincides with usage in the related testing procedures. (Strictly, there should be a notational distinction between a confidence procedure R as a set-valued random variable and a confidence region R as the realization for a particular sample. However, the meaning of the symbol R should be clear from the context or the verbal distinction between confidence procedure and confidence region.)

This method does not necessarily give usable confidence regions, even when the test of each H_λ separately is satisfactory. In order that the tests need not actually be carried out for each λ, some continuity in λ must be required of the family of tests. Any (non-randomized) test of H_λ can be represented (not uniquely) by a statistic $h(\lambda)$ where, for any sample z, H_λ is rejected if and only if $h(\lambda, z) > 0$. If there is a choice of h which is, for a fixed sample z, continuous in λ, then the confidence region R is a closed set with boundary equation $h(\lambda, z) = 0$.

However, continuity of $h(\lambda)$ is not generally enough. If λ is one dimensional; a useful confidence region is usually an interval. A solution providing the limits of the interval is satisfactory, but one providing only a complex equation $h(\lambda) = 0$ for the limits may not be.

When λ is multidimensional, the problems of computation and representation are greatly magnified. The boundary equation will likely increase in complexity rapidly with increasing dimension. But more serious is the difficulty of representing the region even when $h(\lambda)$ is given explicitly in terms of simple functions. The boundary can be plotted in two dimensions, as can cross sections in more than two dimensions, though with effectiveness decreasing with increasing dimension. A principal difficulty is that few shapes are readily visualized in more than two dimensions, or, what is more essential, that comprehension of a region from the equation of its boundary is restricted to very simple surfaces.

The simplest regions are the parallelepipeds which can be completely described by giving limits on each coordinate of a coordinate system related by an affine transformation to the original coordinate system, or equivalently, by giving p linear double inequalities on the coordinates of λ.

The next simplest regions would seem to be the convex polyhedra. When the number of faces is small, the region is simply described by giving the linear inequalities corresponding to each face and is only slightly more complex than the parallelepiped. As an approximate representation of a region with corners, the number of faces is likely too large to permit use of the inequalities and the region must be thought of, with greater difficulty and less adequately, in terms of the corners (vertices).

Ellipsoidal confidence regions are important, largely because they occur naturally in the classical normal theory of means and regression coefficients and also in the general large sample confidence theory. They are probably visualized as rounded boxes and their description by a center and lengths and directions of principal axes corresponds closely to the parallelepiped description.

3. Geometrical idea of intersection confidence regions. In many multidimensional confidence problems, interest centers more on the separate coordinates

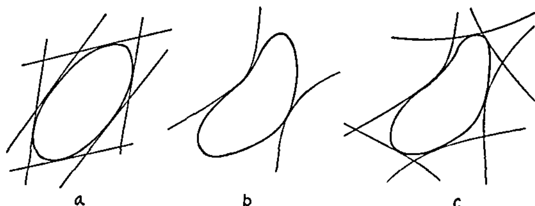


FIG. 1. (a) An ellipse as intersection of straight strips. (b) A standard region and a curved strip. (c) A standard region as intersection of curved strips

or on linear combinations of them than on the multidimensional quantity (e.g., the means in an analysis of variance). The usual ellipsoidal confidence region is of little value. Scheffé's [12] multiple comparisons procedure amounts to representing the ellipsoid as the intersection of all the slabs between parallel pairs of tangent hyperplanes (see figure 1a). Each slab gives a confidence interval for a single linear combination of coordinates. The totality of such intervals has the same joint error level as does the ellipsoidal region. The procedure permits making as many confidence statements on linear combinations as desired and permits the posterior selection of "most interesting" statements.

The same representation by slabs is valid and useful for any convex region and this is the basis for the multiple comparisons procedure given by Tukey [13], Roy and Bose [11] and others.

Even in problems in which the multidimensional quantity is of principal interest, the multiple comparison methods provide a means for approximating convex confidence regions by convex polyhedra, regions more easily described and visualized. Often, the linear inequalities defining the polyhedron are much more easily obtained (computationally) than is the boundary equation of the exact region.

In the small sample theory of more complicated problems (such as the location of a regression surface maximum), the standard confidence regions are not ellipsoids and may not even be convex, connected, or bounded. There is no practical way to determine from a particular boundary equation if the region is convex. The intersection region procedures developed here are an attempt to construct some usable approximate representations for some of these problems.

The idea is to approximate a standard region as the intersection of several regions each of which is fairly easy to represent and to compute. They are typically (in two dimensions) curved strips rather than straight strips (Figures 1b and 1c). The regions are determined essentially by applying the multiple comparisons theory at an earlier stage in the confidence region construction. The approximation is carried one stage further in which the curved strips are approximated by straight strips and their intersections by convex polyhedra.

4. Intersection region procedures for families of general linear hypotheses. The most important class of quantities amenable to intersection region procedures arise from general linear hypotheses in general linear models (cf. Wilks [14], Chap. 8). An n dimensional vector $\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is observed in which $\mathbf{X}$ is a known $n \times m$ matrix of rank m, $\boldsymbol{\beta}$ the unknown m-dimensional vector of 'regression' coefficients and the n components of $\boldsymbol{\varepsilon}$ are independently and normally distributed each with zero mean and variance σ^2. Least squares estimates of $\boldsymbol{\beta}$ and σ^2 are

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{z})$$

$$s^2 = \frac{1}{n-m}[\mathbf{z}'\mathbf{z} - (\mathbf{z}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{z})],$$

distributed independently as a normal with mean $\boldsymbol{\beta}$ and covariance matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ and as $\sigma^2\chi^2/(n-m)$ on $\nu = n - m$ degrees of freedom.

Many quantities of interest can be represented as the roots of sets of simultaneous linear equations in the regression coefficients; as the root in λ of the equations

$$\sum_{j=1}^{m} \beta_j \delta_{ij}(\lambda) = \delta_{i0}(\lambda); \qquad (i = 1, \cdots, p)$$

linear in the regression coefficients, but of arbitrary though specified form in λ.

Linear combinations of means or regression coefficients are included by choosing all $\delta_{ij}(\lambda)$ to be constants and taking $\delta_{i0}(\lambda) = \lambda_i$. Two one-dimensional quantities typical of the more complicated problems motivating the intersection procedures are:

(i) β_1/β_2: $\delta_{11} = 1$, $\delta_{12} = -\lambda_1$, other δ_{ij} and δ_{i0} zero.

(ii) location of vertex of regression curve $\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$: $\delta_{11} = 1$ $\delta_{12} = 2\lambda_1$, $\delta_{13} = 3\lambda_1^2$, other δ_{ij} and δ_{i0} zero.

If λ is the true value of the quantity, it satisfies the equations:

$$H_\lambda : \sum \beta_j \delta_{ij}(\lambda) = \delta_{i0}(\lambda); \qquad (i = 1, \cdots, p)$$

or written in vector form (with natural definitions):

$$H_\lambda : \boldsymbol{\Delta}_\lambda \boldsymbol{\beta} = \boldsymbol{\delta}_{0\lambda}$$

and is a "general linear hypothesis." Any procedure for testing H_λ for every λ leads to a confidence region procedure for λ. Several procedures will be used.

Let $\delta_i(\lambda) = \sum \beta_j \delta_{ij}(\lambda) - \delta_{i0}(\lambda)$ and $\boldsymbol{\delta}_\lambda = [\delta_1(\lambda), \cdots, \delta_p(\lambda)]' = \boldsymbol{\Delta}_\lambda\boldsymbol{\beta} - \boldsymbol{\delta}_{0\lambda}$. For each λ, the least squares estimate of $\boldsymbol{\delta}_\lambda$ is $\mathbf{d}_\lambda = \boldsymbol{\Delta}_\lambda \mathbf{b} - \boldsymbol{\delta}_{0\lambda}$. $\mathbf{d}_\lambda$ is normally distributed with $E(\mathbf{d}_\lambda) = \boldsymbol{\delta}_\lambda$ and Cov $(\mathbf{d}_\lambda) = \sigma^2\boldsymbol{\Delta}_\lambda(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\Delta}_\lambda' = \sigma^2\mathbf{V}_\lambda$. (Note that $\mathbf{d}_\lambda$ and $\mathbf{V}_\lambda$ will generally depend on λ except when $\boldsymbol{\Delta}_\lambda$ is a matrix of constants—as it is for the usual simple problems).

In all that follows, the observations are used only to compute $\mathbf{d}_\lambda$ and $\mathbf{V}_\lambda$ and the sample variance s^2. $\mathbf{V}_\lambda$ *is assumed nonsingular (and hence positive definite) for every* λ.

The likelihood ratio test of the hypothesis $H_\lambda : \delta_\lambda = 0$ is: reject H_λ if $T_\lambda \geq A$ with test statistic

$$T_\lambda = \frac{\mathbf{d}_\lambda' \mathbf{V}_\lambda^{-1} \mathbf{d}_\lambda}{ps^2}.$$

When H_λ is true, T_λ has an F distribution with p and ν degrees of freedom. In general, T_λ has a noncentral F distribution with noncentrality parameter $\delta_\lambda' \mathbf{V}_\lambda^{-1} \delta_\lambda / \sigma^2$. The test with critical value $F_{p,\nu,\alpha}$ (the upper 100 α percent point of the F distribution) is a similar level α test and is the uniformly most powerful invariant level α test of H_λ. Throughout the paper, this test will be called the "standard" test of the general linear hypothesis.

The corresponding confidence procedure for λ is the "standard" level α procedure: $R_S = \{\lambda : T_\lambda \leq F_{p,\nu,\alpha}\}$ which can be written as

$$R_S = \left\{\lambda : \begin{vmatrix} ps^2 F_{p,\nu,\alpha} & \mathbf{d}_\lambda' \\ \mathbf{d}_\lambda & \mathbf{V}_\lambda \end{vmatrix} \geq 0\right\}$$

or

$$R_S = \{\lambda : \sum_{i,j} d_{\lambda i} d_{\lambda j} V_{\lambda ij} - ps^2 F_{p,\nu,\alpha} |\mathbf{V}_\lambda| \leq 0\} \tag{4.1}$$

in which $V_{\lambda ij}$ is the cofactor of the element $v_{\lambda ij}$ in $\mathbf{V}_\lambda$. This confidence procedure has been constructed and used by Box and Hunter [2].

The confidence procedure can be difficult to use especially when the elements of $\mathbf{V}_\lambda$ depend on λ. For then the boundary equation and sometimes the region itself can be very complicated and the necessary computation messy.

The intersection region procedure is based on working separately with the p single equations $\delta_i(\lambda) = 0$ composing H_λ or, more conveniently, with linear combinations of these equations. When $\delta_i(\lambda) = \beta_i - \lambda_i$ so that the quantity of interest is the vector of regression coefficients, the procedure reduces to the multiple comparisons procedure of setting confidence limits on some or all linear combinations of the $\{\beta_i\}$.

Let $\mathbf{k}_1, \cdots, \mathbf{k}_r$ be any r prescribed p-dimensional vectors, and let $H_{\lambda i}$ denote the hypothesis $\mathbf{k}_i' \delta_\lambda = 0$. Every $H_{\lambda i}$ is true when H_λ is, and if the vectors $\{\mathbf{k}_i\}$ span p-space, the truth of all r "component" hypotheses implies that H_λ is true.

Suppose each hypothesis $H_{\lambda i}$ were tested according to: reject $H_{\lambda i}$ if $T_{\lambda i} > A$. A natural joint test of H_λ is to reject H_λ if any $H_{\lambda i}$ is rejected, i.e. if

$$U_\lambda = \max_{1 \leq i \leq r} T_{\lambda i} > A. \tag{4.2}$$

Corresponding to each component set of tests is a confidence region R_i such that

$$R_i = \{\lambda : T_{\lambda i} < A\}$$

and the intersection region R_I is defined as the intersection of the r regions $\{R_i\}$, or equivalently as the region defined by the joint test:

$$R_I = \bigcap_{i=1}^{r} R_i = \{\lambda : U_\lambda < A\}.$$

Each $H_{\lambda i}$ is a linear hypothesis with standard test statistic (which will be used)

$$T_{\lambda i} = \frac{(\mathbf{k}_i' \, \mathbf{d}_\lambda)^2}{s^2(\mathbf{k}_i' \, \mathbf{V}_\lambda \, \mathbf{k}_i)}.$$

The component region R_i is given by

$$R_i = \{\lambda : (\mathbf{k}_i'\mathbf{d}_\lambda)^2 - As^2(\mathbf{k}_i'\mathbf{V}_\lambda\mathbf{k}_i) \leqq 0\}$$

and is usually much simpler, computationally and geometrically, than the standard region R_S. When H_λ is true, each $T_{\lambda i}$ is distributed as $F_{1,\nu}$. The joint distribution of the $T_{\lambda i}$ follows, in principle, from the fact that

$$x_i = \frac{\mathbf{k}_i' \, \mathbf{d}_\lambda}{(\mathbf{k}_i' \, \mathbf{V}_\lambda \, \mathbf{k}_i)^{1/2}}; \qquad (i = 1, \cdots, r) \tag{4.3}$$

are distributed as a multivariate normal with zero means, (when H_λ is true), variances σ^2, and correlations depending on the $\{\mathbf{k}_i\}$ and on λ.

The choice of the critical value A must be a compromise between control of the error level of procedure, ease of computation, and simplicity of the resulting boundary equation. In order that the intersection region procedure have a constant error level α, A must be the 100 α percent point of the distribution of U_λ, the studentized maximum of the squares of correlated normal deviates. But except for a few special cases, these percent points cannot now be obtained without major computation. And since they would likely depend on λ through the correlations, the boundary equations of the intersection region would be complicated by the presence of the function $A(\lambda)$. The use of the exact percent point, if obtainable, for some single "compromise" value of λ might be an excellent choice. I assume throughout that a constant (in λ) critical value is used. Attention here will mainly be restricted to two approximate choices, each "conservative" in the sense that the error level of the intersection region procedure does not exceed the nominal level α.

THEOREM 4.1. *For any set of prescribed* $\mathbf{k}_1, \cdots, \mathbf{k}_r$, *the confidence region procedure* R_I *using critical value* $A = F_{1,\nu;\alpha/r}$ *has error level not exceeding* α.

Since U_λ exceeds A if and only if at least one $T_{\lambda i}$ exceeds A,

$$P(U_\lambda > A) \leqq \sum_{i=1}^{r} P(T_{\lambda i} > A).$$

(This holds generally, without regard for the meaning of $T_{\lambda i}$ and leads to an immediate generalization of the theorem for joint tests and intersection procedures based on any separate tests of any set of component hypotheses.) When H_λ is true, every $T_{\lambda i}$ has an $F_{1,\nu}$ distribution so with $A = F_{1,\nu;\alpha/r}$, the right hand

side is exactly α and the joint test of H_λ has error level not exceeding α. This holds for every λ, so the error level of the intersection region procedure is also so bounded.

The actual error level using $A = F_{1,\nu,\alpha/r}$ will depend on λ through the correlations of the $\{\mathbf{k}_i'\mathbf{d}_\lambda\}$. When these correlations are small, the error level will be quite close to α. Some results on the closeness of the bound α are given in section eight.

As the number r of linear combinations is increased, the correlations increase and the bound gets worse. The behavior of the intersection region is best studied in the limiting case where all linear combinations are used. The distribution theory is exactly that used by Scheffé [12] and is based on an algebraic lemma.

LEMMA (Scheffé). *If* $\mathbf{d}$ *is any p-vector,* $\mathbf{V}$ *any symmetric positive definite* $p \times p$ *matrix, then*

$$\sup_{\text{all } \mathbf{k}} \frac{(\mathbf{k}'\mathbf{d})^2}{(\mathbf{k}'\mathbf{V}\mathbf{k})} = \mathbf{d}'\mathbf{V}^{-1}\mathbf{d}.$$

THEOREM 4.2. *The intersection region procedure using all linear combinations, each with critical value* $A = pF_{p,\nu,\alpha}$, *is identical to the level* α *standard region procedure.*

COROLLARY. *Any intersection region based on r prescribed combinations* $\{\mathbf{k}_i\}$ *and critical value* A_0 *always contains the standard region with error level*

$$P\{F_{p,\nu} > A_0/p\}.$$

Applying the lemma to $\mathbf{d}_\lambda$ and its covariance matrix $\sigma^2\mathbf{V}_\lambda$ and studentizing with s^2,

$$\sup_{\text{all } \mathbf{k}} \frac{(\mathbf{k}'\mathbf{d}_\lambda)^2}{s^2(\mathbf{k}'\mathbf{V}_\lambda \mathbf{k})} = \frac{\mathbf{d}_\lambda' \mathbf{V}_\lambda^{-1}\mathbf{d}_\lambda}{s^2}.$$

But the left-hand side is the test statistic associated with the intersection procedure (over all $\mathbf{k}$) and the right-hand side is p times the standard test statistic.

Any intersection procedure can be treated as an approximation to some standard procedure. The intersection region will always contain the standard region and will converge to it as more linear combinations are used. The gain in simplicity of the component regions may more than compensate for the large number of regions and the imperfect approximation.

The use of $A = pF_{p,\nu,\alpha}$ has one advantage over all other choices, approximate or exact, for a finite r. The distribution theory of the $\{T_{\lambda i}\}$ and related statistics is valid only if the vectors $\{\mathbf{k}_i\}$ are chosen independently of $\mathbf{d}_\lambda$ and s^2. But Theorem 4.2 is based on all linear combinations and thus can be used for $\{\mathbf{k}_i\}$ selected after studying the data. A useful a posteriori choice of linear combinations will be illustrated in the application in section seven. (This advantage is a primary motivation of Scheffé's [12] multiple comparison procedure.)

When the equations defining λ are homogeneous linear functions of the regression coefficients, confidence regions for λ can have shapes and behavior not

occurring in classical confidence regions. Several such properties follow from Theorem 4.3.

THEOREM 4.3. *If for all* λ, *each component of* δ_λ *is a homogeneous linear function of the regression coefficients of the linear model, then for any sample there is a nonzero* α^* *which will depend on the sample but not on* λ, *such that the standard confidence region for* λ *is the entire space for any error level less than* α^*.

COROLLARY. *The theorem holds, with a possibly different* α^*, *for the intersection confidence region.*

The theorem follows easily from the well-known interpretation of the test statistic T_λ of the general linear hypothesis, that

$$T_\lambda = \frac{S_{1\lambda} - S_0}{S_0} \cdot \frac{\nu}{p}$$

where $S_0 = \nu s^2$ is the sum of squares of residuals after the least squares fit of the linear model to the data, while $S_{1\lambda}$ is the sum of squares of residuals after the best least squares fit subject to the restriction of the hypothesis $H_\lambda : \delta_\lambda = 0$ (cf. Wilks [14]). One of the possible fits satisfying the homogeneous restrictions $\delta_\lambda = 0$ is that with all regression coefficients estimated to be zero, leaving a residual sum of squares $\sum z^2$, the original sum of squares. Consequently, $S_{1\lambda} \leqq \sum z^2$ for all λ and for any sample, T_λ is bounded by a constant depending on the sample but not on λ. Since $F_{p,\nu;\alpha}$ approaches infinity as α approaches zero, for the sample z there is a nonzero $\alpha^*(z)$ for which H_λ will be accepted for all λ at any significance level $\alpha < \alpha^*(z)$, and the corresponding confidence region will be the entire space. The corollary follows using the corollary to Theorem 4.2.

The theorem shows that the confidence regions need not be bounded. Since the value(s) of λ for which T_λ is maximum will generally be finite, the confidence region as a function of the error level will close in around the maximizing point(s), and the resulting region will be neither convex nor simply connected and perhaps not even connected. (In the usual problems with means and regression coefficients, the hypotheses are not homogeneous and, what is essential, the constant term depends on λ.)

5. Geometry of intersection regions for a class of equations linear in λ. Further study of intersection regions requires specifying the form in λ of the defining equations. An interesting class of equations is suggested by the problem of locating the maximum of a quadratic regression surface (section seven). Suppose that λ is a p-dimensional vector $\boldsymbol{\lambda}$ and that the equation $\delta_\lambda = 0$ is linear and *homogeneous* in the regression coefficients and *linear* in $\boldsymbol{\lambda}$. Introduce the notation $\delta_\lambda = \gamma + \boldsymbol{\Gamma}\boldsymbol{\lambda}$ in which the elements of the p-vector γ and the $p \times p$ matrix $\boldsymbol{\Gamma}$ are homogeneous linear functions of the regression coefficients. Let $\mathbf{c}$ and $\mathbf{C}$ be the corresponding least squares estimates of γ and $\boldsymbol{\Gamma}$ and let $\mathbf{d}_\lambda = \mathbf{c} + \mathbf{C}\boldsymbol{\lambda}$. The covariance matrix $\sigma^2\mathbf{V}_\lambda$ of $\mathbf{d}_\lambda$ is an inhomogeneous quadratic function of $\boldsymbol{\lambda}$. The particular forms of γ and $\boldsymbol{\Gamma}$ are of no interest except for the evaluation of $\mathbf{d}_\lambda$ and $\mathbf{V}_\lambda$ (a tedious but straightforward task) and to verify the two

assumptions: *Assume that* $\mathbf{V}_\lambda$ *is nonsingular for all* λ. *Assume that the (random) matrix* $\mathbf{C}$ *is nonsingular with probability one.*

Then a unique solution $\hat{\lambda}$ (the maximum likelihood estimate of λ) of $\mathbf{d}_\lambda = 0$ will exist. $\mathbf{\Gamma}$ may be singular for a particular set of population regression coefficients, so that the "population value of λ" need not be unique or even exist. (If no solution of $\boldsymbol{\delta}_\lambda = 0$ exists, the confidence problem is vacuous.)

The development of this section concerns geometric properties of intersection regions. Throughout, the observed sample will be held fixed, arbitrarily, except for the above mentioned set of probability zero.

The component region R_i based on the combination $\mathbf{k}_i'\boldsymbol{\delta}_\lambda$ is $\{\lambda: h_i(\lambda) \leqq 0\}$ with quadratic boundary equation $h_i(\lambda) = (\mathbf{k}_i'\mathbf{d}_\lambda)^2 - As^2(\mathbf{k}_i'\mathbf{V}_\lambda\mathbf{k}_i) = 0$. Since $\mathbf{V}_\lambda$ is positive definite for all λ, all points on the hyperplane $\mathbf{k}_i'\mathbf{d}_\lambda = 0$ (call it M_i) lie in the interior of R_i. The two parts (if they exist) of the exterior (complement) of R_i on either side of M_i are each convex (Theorem 9.1).

Thus, R_i is the region between the sheets of a two-sheeted hyperboloid, the exterior of an ellipsoid, or limiting and transitional forms of these. The boundary (call it F_i) can never be one of the one-sheeted hyperboloids.

Considered as a function of the critical value A (or of the significance level), R_i is the hyperplane M_i when $A = 0$ and expands monotonically with A, first as a "curvilinear slab" between the two sheets of a hyperboloid of small curvature, then widening and curving until it eventually becomes the outside of an ellipsoid and finally fills the entire space. By Theorem 4 3, this last will occur for a finite A.

The component region will be most easily described and comprehended when it is a "curvilinear slab," being almost a confidence interval for the linear combination $\mathbf{k}_i'\mathbf{d}_\lambda$. In any case, the computations involved in using R_i are relatively simple.

The component regions are studied as a preliminary to forming intersection regions. In this section, restrict attention to p linearly independent combinations. The intersection of p slabs is a parallelepiped. At best, the component regions are bounded by hyperboloids of small curvature, and the intersection of p such regions is a "curvilinear parallelepiped." The $2p$ "faces" become concave and the 2^p corners are moved at least enough that the 2^{p-1} corners on any "face" are not coplanar.

Let $R_I = \bigcap R_i$ be the intersection region determined by $\mathbf{k}_1, \cdots, \mathbf{k}_p$ and let $J = \bigcap F_i$ be the intersection of the boundaries. The points of J—the "corners" of R_I—are the solutions of the p simultaneous quadratic equations

$$\{h_i(\lambda) = 0, i = 1, \cdots, p\}.$$

Except with probability zero, J will contain not more than 2^p points. Let R^1 be the convex closure of J. R^1 is a convex polyhedron, all of whose vertices are points of J (not necessarily conversely).

In the intuitive discussion above, R_I would be contained in R^1 so that R^1 would be a conservative and perhaps close approximation to R_I. But the ap-

proximation depends critically on R_I being a "curvilinear parallele
R_I is unbounded, disconnected, or otherwise misshaped, the formal c
of R^1 will usually lead to a clearly bad approximation, but it can l
apparently good but erroneous approximation. Complex shaped in
regions may arise because of poor choices of components (e.g., too h
related) or because of the inadequacy of the sample data in accurat
mining λ. The R^1 approximation can be so bad that it has no points i
with R_I except the corners. This will occur if all the corners lie on t
one boundary (say F_i) on one side of its hyperplane M_i, as will necess
if any R_i is an ellipsoid (Theorem 9.2).

A positive result, representing the behavior understood in the int
cussion is that R^1 contains R_I if there are exactly 2^p corners, one in e
2^p parts of p-space formed by the p hyperplanes $\{M_i\}$ (Theorem 9.3)

To use R^1, one must find all corners of R_I. There is need for son
approximations, to provide a first approximation for the complicate
tions of the corners of R_I, and also to be easier to describe and use t
Two simple polyhedral approximations R^2 and R^3 are suggested, both
the idea that the boundaries $\{F_i\}$ are hyperboloids of small curvatur
in the region of interest.

R^2 is obtained by replacing each hyperboloid F_i by a pair of tang
planes and taking the convex polyhedron formed by these $2p$ hyper
is the parallelepiped formed by approximating each F_i by a pair of hy
both parallel to the corresponding M_i. These approximations are ea
mined once the point of tangency or intersection on each boundary
The suggested choice is the pair of points lying also on every M_j,
for each i, the two points lie on opposite sides of M_i (the reverse is a
cation that R_I is not a "curvilinear parallelepiped"), R^2 is always co
R_I (Theorem 9.4). If the hyperboloids are nearly flat, the approximati
By dilating the region until all corners are outside R_I, a "conserv
proximation of the same convenient shape could be obtained. R^2 *or a*
probably the most useful of all the suggested regions. R^3 is a much rough
mation whose virtue is simplicity of shape and computation.

6. On the computation of intersection regions and approximations
tions linear in λ. The computations needed in the approximations
five are simplified by a change of coordinates. For a particular
$\mathbf{k}_1, \cdots, \mathbf{k}_p$, and a particular sample, define new (oblique) coord
the space of λ by

$$\xi_i = \mathbf{k}_i'\mathbf{d}_\lambda ; \qquad (i =$$

The inverse transformation must be obtained, requiring the solution
taneous linear equations in $\{\lambda_i\}$. (The unique solution is guarante
assumed linear independence of the $\{\mathbf{k}_i\}$ and nonsingularity of $\mathbf{C}$.)

The coordinate hyperplane $\xi_i = 0$ is M_i and the maximum likelihoo

$\hat{\lambda}$ is at the origin $\xi = 0$. Each $h_i(\lambda)$ can be written in ξ coordinates as

$$h_i(\lambda) = g_i(\xi) = \xi_i^2 - As^2(k_i'V_\lambda k_i)$$

and is a quadratic function of the $\{\xi_i\}$.

The computation of R^2 and R^3 is immensely simplified. The pair of points lying on F_i and every M_j, $j \neq i$ have all ξ coordinates but the ith zero, and that given by the roots $(e_i' < e_i'')$ of the quadratic $g_i(0, \cdots, 0, \xi_i, 0, \cdots, 0) = 0$. If they have the same sign, the intersection region, at least for the particular choice of components, is dangerously complicated. If $e_i' < 0 < e_i''$, then the equation of the tangent hyperplane approximation to the part of F_i with $\xi_i > 0$ (say) is

$$0 = \sum_{\substack{j=1 \\ j \neq i}}^{p} \xi_j \left[\frac{\partial g_i(\xi)}{\partial \xi_j}\right]_{\xi=(0,\cdots,e_i'',\cdots,0)} + (\xi_i - e_i'')\cdot\left[\frac{\partial g_i(\xi)}{\partial \xi_i}\right]_{\xi=(0,\cdots,e_i'',\cdots,0)}.$$

and is easily written down when $g_i(\xi)$ is given explicitly. R^2 is defined by the $2p$ corresponding inequalities, each chosen so that $\xi = 0$ satisfies it. R^3 is given as $\{\xi: e_i' < \xi_i < e_i''; i = 1, \cdots, p\}$. The inequalities for R^2 or R^3 are easily changed back to linear inequalities in λ if desired.

The R^1, R^2 and R^3 constructions have no unique or natural extensions to more than p component regions (in p dimensions). One possible procedure would be to repeat the R^2 construction for another set of p components (perhaps with some overlaps), using a new set of coordinates, then converting both sets of inequalities to some one convenient coordinate system. The approximate region would then be the convex polyhedron defined by all of the inequalities. Some dilation of the R^2 region as discussed in section five would be desirable to prevent serious underapproximation of the confidence region

7. An application to the location of the maximum of a quadratic regression surface. Many aspects of the problem of determining the values of the input variables of a process to yield a maximum response have been studied by Box and colleagues ([1], [2], [3], [4], [5]). Here we use the simple model in which each observed response z is distributed normally and independently with variance σ^2 and mean

(7.1) $$E(z) = \gamma_0 + \gamma'y + \tfrac{1}{2}y'\Gamma y$$

with input variables $y' = (y_1, \cdots, y_p)$ and regression coefficients

$$\gamma_0, \gamma' = (\gamma_1, \cdots, \gamma_p), \qquad \Gamma = [\gamma_{ij}]$$

with $\gamma_{ij} = \gamma_{ji}$. n sets of (z, y) comprise the observed data and the model is a "general linear model" of section four with $m = 1 + p + p(p+1)/2$ regression coefficients. Denote by c_0, c, C and s^2 the least squares estimates of γ_0, γ, Γ and σ^2. The fitted surface is

(7.2) $$\hat{z} = c_0 + c'y + \tfrac{1}{2}y'Cy.$$

The estimates and their variances and covariances can be computed from the formulas of section four or from those given by Box and Wilson [4].

If the surface (7.1) has a maximum at $\mathbf{y} = \boldsymbol{\lambda}$, $\boldsymbol{\lambda}$ will satisfy the stationarity equation $H_\lambda : \boldsymbol{\gamma} + \boldsymbol{\Gamma\lambda} = 0$. This equation will be satisfied by any vertex—maximum, minimum or saddle point—of the surface and all confidence regions are for the location of a vertex, type unspecified. Box and Hunter [2] construct the standard confidence region and show that if the region is bounded, then the region can be said to represent one particular type of vertex in the sense that for every $\boldsymbol{\lambda}$ in the region and for each fitted surface with vertex $\boldsymbol{\lambda}$ that does not give a "significantly poor fit," the vertex is of the same type. Their argument extends to intersection regions based on joint tests of the form of equation (4.2.).

From the structure of $\mathbf{d}_\lambda = \mathbf{c} + \mathbf{C}\boldsymbol{\lambda}$ it follows that even with the maximum of balance and symmetry in the design, each diagonal element of the covariance matrix $\mathbf{V}_\lambda$ of $\mathbf{d}_\lambda$ has at least a constant term and all p square terms. Each off-diagonal element has at least a term in $\lambda_i\lambda_j$. Consequently, in the equation (4.1) for the standard confidence region R_S, each term will generally be a polynomial in the $\{\lambda_i\}$ of degree $2p$. Even for $p = 2$, the equation is already unwieldly. Box and Hunter [3] show that with a rotatable design, the equation can be reduced to a quartic for any p.

H_λ is exactly of the form studied in section five and the nonsingularity assumptions for $\mathbf{V}_\lambda$ and $\mathbf{C}$ are met. Intersection region procedures are applicable. The simplest choice of linear combinations is the direct use of the p components of $\mathbf{d}_\lambda$. With this choice, the critical value $A = F_{1,\nu;\alpha/p}$ could be used to give an error level bounded by α.

A better choice is suggested by a canonical analysis of the fitted surface such as is obtained by introducing new coordinates $\{x_i\}$ with origin at the center and with axes the principal axes of the fitted surface (7.2). If $\{\mathbf{m}_i\}$ and $\{b_{ii}\}$ are the eigenvectors and corresponding eigenvalues of $\mathbf{C}$ and if $b_i = \mathbf{m}_i'\mathbf{c}$, then $x_i = \mathbf{m}_i'\mathbf{y} + b_i/b_{ii}$ and the fitted surface becomes $\hat{z} = \text{constant} + \frac{1}{2}\sum_{i=1}^{p} b_{ii}x_i^2$. (Box and Wilson [4] and Box [1] give more details and interpretations.)

The suggested choice for $\mathbf{k}_i$ is $\mathbf{m}_i/b_{ii}$ which corresponds to using the separate stationarity equations found by equating to zero the partial derivatives of the true surface in the directions of the principal axes of the fitted surface. This set being dependent on the data, the only simple valid critical value is $A = pF_{p,\nu;\alpha}$ (with which arbitrarily many more combinations can be used without increasing the error level of the intersection region). With $\mathbf{k}_i = \mathbf{m}_i/b_{ii}$, the linear form defining ξ_i and used for the component region R_i is

$$\begin{aligned} \xi_i &= \mathbf{k}_i'\mathbf{d}_\lambda = \mathbf{m}_i'\mathbf{d}_\lambda/b_{ii} = \mathbf{m}_i'\mathbf{c}/b_{ii} + \mathbf{m}_i'\mathbf{C}\boldsymbol{\lambda}/b_{ii} \\ &= b_i/b_{ii} + \mathbf{m}_i'\boldsymbol{\lambda}. \end{aligned} \tag{7.3}$$

Thus, the transformation to the ξ coordinate system, useful in the computation of intersection region approximations, is here identical to the transformation to the principal coordinates $\{x_i\}$ useful in understanding the shape of the regression surface.

The intersection region R_I and the approximation R^1, R^2 and R^3 will be illustrated on the numerical example given by Box and Hunter [2] to illustrate R_S. For the fitted surface $c_0 = 77.95$, $s^2 = 1.07$, $\nu = 9$,

$$\mathbf{c} = \begin{bmatrix} 3.76 \\ -1.57 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} -5.74 & 3.84 \\ 3.84 & -5.28 \end{bmatrix}$$

with covariance matrix given by equation (27) of [2]. The fitted surface has a maximum $\hat{\lambda} = (.889, .349)$. The surface in the *principal coordinates* x *is*: $\hat{z} = 79.35 - \frac{1}{2}(9.357\, x_1^2 + 1.663\, x_2^2)$ and the transformation of coordinates is

$$(7.4) \qquad \begin{aligned} x_1 &= .7279\, y_1 - .6857\, y_2 - .4076 \\ x_2 &= .6857\, y_1 + .7279\, y_2 - .8631. \end{aligned}$$

The transformation consists of a translation of the point $\hat{\lambda}$ to the origin and a rotation through $-43°17.2'$.

Using the second set of recommended $\{\mathbf{k}_i\}$ and the critical value $A = 2F_{2,9,.05} = 8.52$, the transformation (7.3) from the λ to ξ coordinates is identical with the transformation (7.4) from $\mathbf{y}$ to $\mathbf{x}$ coordinates. The boundary equations in the ξ coordinates of the two component regions R_1 and R_2 are:

$$g_1(\xi) = -.035 - .014\, \xi_1 - .054\, \xi_2 + .929\, \xi_1^2 - .045\, \xi_2^2 + .035\, \xi_1\xi_2$$

$$g_2(\xi) = -.433 + .048\, \xi_1 - .468\, \xi_2 - 1.414\, \xi_1^2 + .514\, \xi_2^2 + .709\, \xi_1\xi_2$$

The corners of the intersection region R_I are $(-.88, 3.10)$, $(.43, 1.31)$, $(-.12, -.55)$, $(.16, -.66)$ of which one lies in each quadrant, satisfying the hypothesis of Theorem 9.3, so that R_I is contained in the quadrilateral R^1 *formed as convex closure* of these points. The parallelepiped approximation R^3 is given by the inequalities $(-.19 \leqq \xi_1 \leqq .20)$, $(-.57 \leqq \xi_2 \leqq 1.48)$ and the polyhedral approximation R^2 is the intersection of the four tangent half spaces:

$$T_1^- = \{\xi: .067 + .360\, \xi_1 + .061\, \xi_2 \geqq 0\}$$

$$T_1^+ = \{\xi: .073 - .360\, \xi_1 + .047\, \xi_2 \geqq 0\}$$

$$T_2^- = \{\xi: .599 + .355\, \xi_1 + 1.053\, \xi_2 \geqq 0\}$$

$$T_2^+ = \{\xi: 1.558 - 1.097\, \xi_1 - 1.053\, \xi_2 \geqq 0\}$$

The regions R_1, R_2, R_I, R^1, R^2, R^3, and R_S (the latter taken from [2], p. 198) are illustrated in Figure 2. (In two dimensions, the approximations to R_I are unnecessary except for simple analytic description and are shown principally to illustrate the different approximations.)

8. Bounds on the error level of intersection procedures. A lower bound on the error level can be obtained that gives some indication of the closeness of the bound in Theorem 4.1.

Fix λ, let $u_i = x_i/s$ with x_i defined by equation (4.3) and let ρ_{ij} = correlation (x_i, x_j). The joint distribution under H_λ of the $\{u_i\}$ is an r-variate generalization

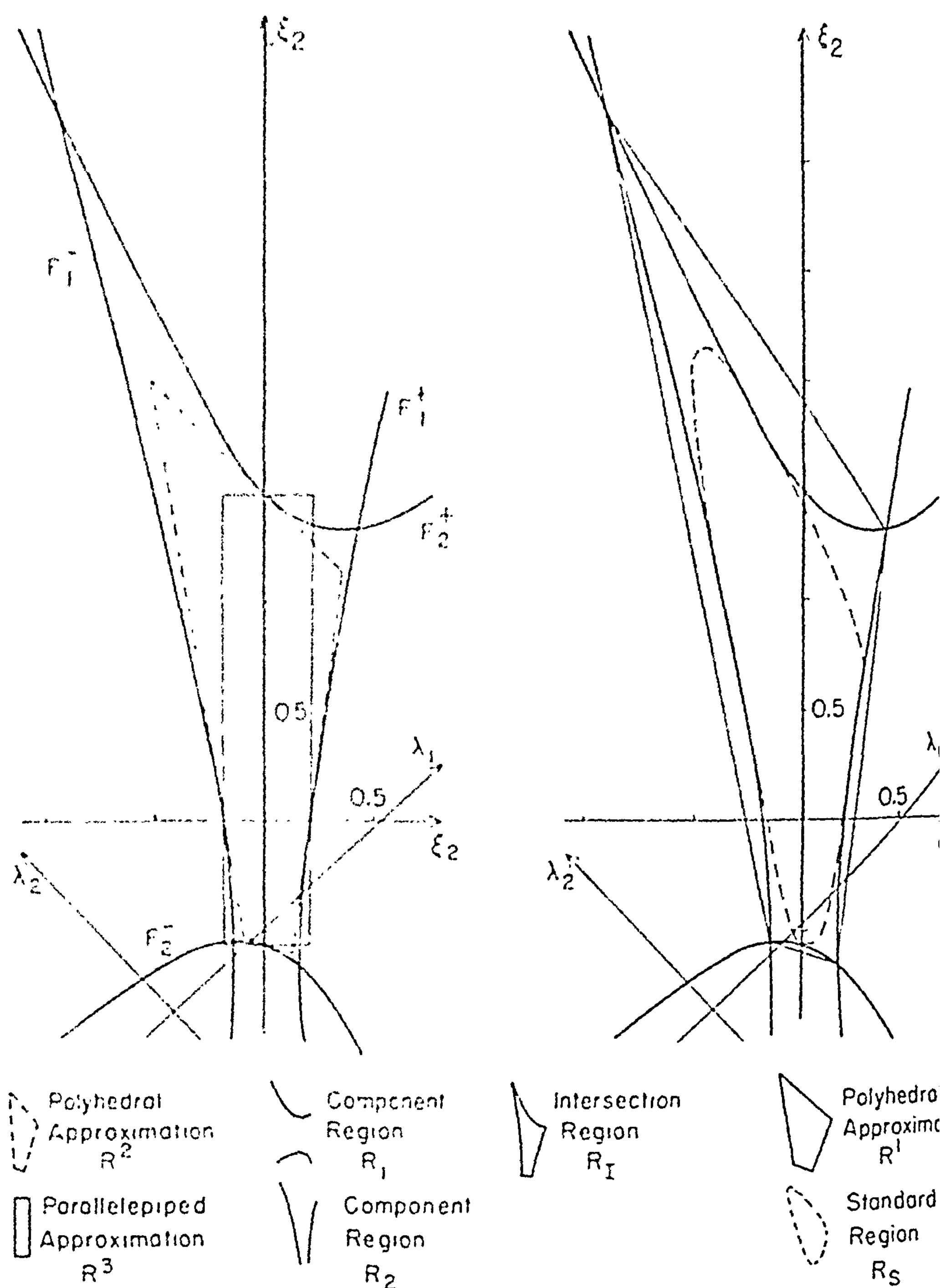


FIG. 2. Various approximate confidence regions for the location of the maxim[um for] the quadratic regression example of section seven.

of the t distribution (cf. Dunnett and Sobel [6]) with ν degrees of freedo[m and] correlation matrix $[\rho_{ij}]$ of the associated r-variate normal distribution. D[enote] the bivariate distribution integrals by

$$d_\nu(a, b, \rho_{ij}) = P(u_i > a, u_j > b)$$

$$f_\nu(a, b, \rho_{ij}) = P(|u_i| > a, |u_j| > b)$$

K. Pearson [10, Table VIII and IX] gives $d_\infty(a, b, \rho)$ for a selection of a, b, ρ and Dunnett and Sobel [6] give something simply related to $d_\nu(a, a, \pm 0.5)$ for a selection of a and ν, and formulas for computing other values. The marginal $f_\nu(a, 0, \rho)$ is the double-tail probability of Student's t distribution and is independent of ρ.

The error level of the joint test is

$$P(U_\lambda > A) = P\left(\max_{1 \leq i \leq r} |u_i| \geqq \sqrt{A}\right.$$

Bounds on this probability are given in Theorem 8.1

THEOREM 8.1.

$$(1) \qquad P\left(\max_{1 \leq i \leq r} |u_i| \geqq a\right) \leqq rf_\nu(a, 0, -)$$

$$(2) \qquad P\left(\max_{1 \leq i \leq r} |u_i| \geqq a\right) \geqq rf_\nu(a, 0, -) - \sum_{i<j} f_\nu(a, a, \rho_{ij})$$

$$(3) \qquad \begin{aligned} P\left(\max_{1 \leq i \leq r} |u_i| \geqq a\right) \geqq{}& rf_\nu(a, 0, -) \\ & - \binom{r}{2}\left[\left(1 - \frac{\rho_1}{\rho_0}\right) f_\nu(a, a, 0) + \frac{\rho_1}{\rho_0} f_\nu(a, a, \rho_0)\right] \end{aligned}$$

in which $\rho_0 = \max |\rho_{ij}|$ *and* $\rho_1 = \sum_{i<j} |\rho_{ij}| \Big/ \binom{r}{2}$.

Equality occurs in (1) *if* $r = 1$ *and in* (2) *and* (3) *if* $r = 2$.

Inequalities (1) and (2) are direct applications of Bonferroni's inequalities (cf. Feller [7]) to the events $\{|u| \geqq a\}$. The inequality (3) follows on combining with inequality (2) the symmetry and convexity of $f_\nu(a, a, \rho)$ proved in the lemma below. For since $0 \leqq |\rho_{ij}| \leqq \rho_0$,

$$f_\nu(a, a, \rho_{ij}) = f_\nu(a, a, |\rho_{ij}|) \leqq \left(1 - \frac{|\rho_{ij}|}{\rho_0}\right) f_\nu(a, a, 0) + \frac{|\rho_{ij}|}{\rho_0} f_\nu(a, a, \rho_0)$$

and

$$\sum_{i<j} f_\nu(a, a, \rho_{ij}) \leqq \binom{r}{2}\left[\left(1 - \frac{\rho_1}{\rho_0}\right) f_\nu(a, a, 0) + \left(\frac{\rho_1}{\rho_0}\right) f_\nu(a, a, \rho_0)\right].$$

LEMMA. $f_\nu(a, a, \rho)$ *is a symmetric, convex function of* ρ.

Use $f_\nu(a, a, \rho) = 2d_\nu(a, a, \rho) + 2d_\nu(a, a, -\rho)$. Write each d_ν as the double integral of the bivariate t-density, change to "elliptical polar coordinates" ([6], p. 154), and integrate out the angle. Straightforward calculation of $\partial^2 f_\nu/\partial\rho^2$ shows convexity.

Table 8.1 gives the upper bound (1) and lower bound (3) for a selection of values of a, r, ν, ρ_0, ρ_1 chosen to give upper bound near .05 or .01. The upper bound would seem to be sufficiently accurate for most uses provided the correlations and r are not very large.

TABLE 8.1

Bounds on $P(\max_{1 \leq i \leq r} |u_i| > a)$ *from Theorem* 8.1

r	ν	a	Upper Bound (1)	Lower Bound (3)					
				$\rho_0 = 0$, $\rho_1 = 0$	$\rho_0 = .5$		$\rho_0 = .8$		
					$\rho_1 = .2$	$\rho_1 = .5$	$\rho_1 = .2$	$\rho_1 = .5$	$\rho_1 = .8$
2	∞	2.2	.056	.055*	—	.052*	—	—	.045*
3	∞	2.4	.049	.048*	.046	.043	—	.038	.031
5	∞	2.6	.047	.046*	.043	.038	.038	.027	.016
8	∞	2.7	.055	.054*	.048	.039	.039	.016	<0
10	∞	2.8	.051	.050*	.043	.033	.033	.007	<0
2	∞	2.8	.0102	.0102*	—	.0098*	—	—	.0080*
3	∞	2.9	.0112	.0112*	.0109	.0104	—	.0092	.0080
5	∞	3.1	.0097	.0096*	.0092	.0086	.0084	.0065	.0046
8	∞	3.2	.0110	.0109*	.0102	.0092	.0086	.0050	.0014
10	∞	3.3	.0097	.0096*	.0089	.0079	.0071	.0033	<0
2	14	2.5	.051	.049*	—	.046*			
5	14	3.0	.048	.044	.039	.033			
3	8	3.0	.051	.047	.044	.040			
2	6	3.0	.048	.045*	—	.042*			
5	7	3.5	.050	.041	.036	.029			

* Exact value.
—Impossible ρ_0, ρ_1 combination.

9. Mathematical results for the geometry of section five. The notation of sections five and six will be used, and all results are for a fixed sample. For terminological convenience, all work will be done in terms of a Euclidean p-space E_p with *rectangular* coordinates ξ. The affine transformation of the space does not affect the properties of interest.

Let Var $\xi_i = \sigma^2(\mathbf{k}_i'\mathbf{V}_\lambda\mathbf{k}_i)$ be the variance of the linear form defining ξ_i, transformed for the particular sample to a function of ξ. If S_i is any set in E_p indexed by i, let

$$S_i^+ = S_i \cap \{\xi\colon \xi_i > 0\}, \qquad S_i^- = S_i \cap \{\xi\colon \xi_i < 0\}, \qquad S_i^0 = S_i \cap \{\xi\colon \xi_i = 0\}.$$

Let $S^* = E_p - S$, and $\bar{S} =$ closure of S.

THEOREM 9.1. *R_i^{*+} and R_i^{*-} are convex.*

To prove R_i^{*+} convex, it is sufficient to show that for any two points ξ_1 and ξ_2 in R_i^{*+} and any constant θ such that $0 < \theta < 1$, $\xi_0 = \theta\xi_1 + (1 - \theta)\xi_2$ is in R_i^{*+}. But

$$\xi \,\varepsilon\, R_i^{*+} \Leftrightarrow \{g_i(\xi) > 0 \quad \text{and} \quad \xi_i > 0\}.$$

$\xi_{0i} > 0$ is immediate. Expanding $g_i(\xi_0)$, using the Cauchy inequality for covariance, $g_i(\xi_j) > 0$ for $j = 1, 2$ and $\xi_{1i}\xi_{2i} > 0$,

$$\begin{aligned} g_i(\xi_0) &= \theta^2[\xi_{1i} - (As^2/\sigma^2)\operatorname{Var}\xi_{1i}] \\ &+ (1-\theta)^2[\xi_{2i} - (As^2/\sigma^2)\operatorname{Var}\xi_{2i}] \\ &+ 2\theta(1-\theta)[\xi_{1i}\xi_{2i} - (As^2/\sigma^2)\operatorname{Cov}(\xi_{1i}, \xi_{2i})] \\ &\geqq \theta^2 g_i(\xi_1) + (1-\theta)^2 g_i(\xi_2) \\ &+ 2\theta(1-\theta)[\xi_{1i}\xi_{2i} - (As^2/\sigma^2)\sqrt{(\operatorname{Var}\xi_{1i})(\operatorname{Var}\xi_{2i})}] \\ &> 2\theta(1-\theta)[\xi_{1i}\xi_{2i} - (\xi_{1i}^2\xi_{2i}^2)^{1/2}] = 0. \end{aligned}$$

COROLLARY. *$\overline{R_i^{*+}}$ and $\overline{R_i^{*-}}$ are convex.*

THEOREM 9.2. *If for some i and some sign (say $+$), $J \cap F_i^+ = J$, then $R_I \cap R^1 \subset F_i^+$.*

By hypothesis, $J \subset F_i^+ \subset \overline{R_i^{*+}}$. By the corollary to Theorem 9.1, $\overline{R_i^{*+}}$ is convex and closed so that $R^1 \subset \overline{R_i^{*+}}$.

$$R_I \cap R^1 \subset R_I \cap \overline{R_i^{*+}} \subset R_I \cap (R_i^{*+} \cup F_i^+) \subset R_I \cap F_i^+ \subset F_i^+.$$

COROLLARY. *If any boundary F_i is an ellipsoid, $R_I \cap R^1 \subset F_i$.*

The entire ellipsoid must be on one side of the plane $\xi_i = 0$.

THEOREM 9.3. *If J contains exactly 2^p points, with one point in each of the 2^p open orthants formed by the p coordinate hyperplanes $M_1, \cdots, M_p$, then $R_I \subset R^1$.*

Three lemmas will be proved, from which the theorem follows immediately.

By hypothesis, J contains 2^p points, one in each open orthant defined by the p coordinate planes $\{\xi_i = 0\}$. Denote the 2^p points ("corners") by

$$\{e_u' = (e_{u1}, \cdots, e_{up}); u = 0, \cdots, 2^p - 1\}$$

with the subscript u assigned so that if $[u_1, \cdots, u_p]$ is the binary expansion of u, then $e_{uj} > 0$ ($e_u \in M_j^{*+}$) if $u_j = 1$ and $e_{uj} < 0$($e_u \in M_j^{*-}$) if $u_j = 0$. The point e_u is in the diagonally opposite orthant from e_{2^p-1-u}. The "diagonal line" $D(u, 2^p - 1 - u)$ through e_u and e_{2^p-1-u} can be parametrized as

$$\{\xi: \xi = \theta e_u + (1-\theta)e_{2^p-1-u}\}$$

and is the union of three disjoint parts: the "diagonal segment" $D^0(u, 2^p - 1 - u)$ with $0 \leqq \theta \leqq 1$, and two "outer diagonals" $D(u)$ with $\theta > 1$ and $D(2^p - 1 - u)$ with $\theta < 0$. $D(u)$ is contained in the same open orthant as e_u. Define

$$Q_i^+ = \text{convex hull of} \bigcup_{\{u: u_i=1\}} D(u)$$

$$Q_i^- = \text{convex hull of} \bigcup_{\{u: u_i=0\}} D(u)$$

$$Q = \bigcup_{i=1}^{p} (Q_i^+ \cup Q_i^-).$$

LEMMA 1. *Under the conditions of Theorem 9.3, $R_I \subset Q^*$.*

For each value of u and i, the diagonal $D(u, 2^p - 1 - u)$ intersects the boundary F_i in the two points e_u and e_{2^p-1-u}. The diagonal segment crosses each

coordinate hyperplane M_i. Since the boundary equation of R_i is quadratic, the segment $D^0(u, 2^p - 1 - u)$ is in R_i and the outer diagonal D_u is in R_i^*. Then

$$\bigcup_{\{u:u_i=1\}} D_u \subset R_i^{*+} \quad \text{and} \quad \bigcup_{\{u:u_i=0\}} D_u \subset R_i^{*-}.$$

By Theorem 9.1, R_i^{*+} and R_i^{*-} are convex, so $Q_i^+ \subset R_i^{*+} \subset R_i^*$ and $Q_i^- \subset R_i^{*-} \subset R_i^*$ for every i. Then $Q \subset \bigcup_{i=1}^p R_i^* = R_I^*$ and $R_I \subset Q^*$.

LEMMA 2. *Under the conditions of Theorem 9.3, R^1 contains a cube with the origin in the interior.*

By induction on the dimension p, the cube with faces $|\xi_i| = a = \min_{u,j}|e_{uj}|$ will be shown to be in R^1. Let ξ^0 be any point with $|\xi_i^0| \leqq a$ for all i. If $p = 1$, then $e_{01} \leqq -a \leqq \xi_1^0 \leqq a \leqq e_{11}$ for the one coordinate and ξ^0 is in the convex closure of $\mathbf{e}_0$ and $\mathbf{e}_1$.

For arbitrary p, let $\mathbf{e}_u$ be an arbitrary corner. Suppose that $e_{up} > 0$. Let $\mathbf{e}_{u'}$ be the corner with sign $e_{u'j} =$ sign e_{uj} for $j < p$ and $e_{u'p} < 0$. Then $e_{u'p} \leqq -a \leqq \xi_p^0 \leqq a \leqq e_{up}$ and there is a convex linear combination $\mathbf{d}_u = \theta_u \mathbf{e}_u + (1 - \theta_u)\mathbf{e}_{u'}$ with $0 \leqq \theta_u \leqq 1$ such that $d_{up} = \xi_p^0$. For the other coordinates, $|d_{uj}| \geqq \min(|e_{uj}|, |e_{u'j}|) \geqq a$ and sign $d_{uj} =$ sign e_{uj}. There are 2^{p-1} points $\mathbf{d}_u$ with $u = (u_1, \cdots, u_{p-1}, 1)$, satisfying the conditions of the theorem on the $p - 1$ dimensional hyperplane $\xi_p = \xi_p^0$ with restricted $\min_{u,j}|d_{uj}| \geqq a$. By the induction hypothesis, the point ξ^0 lies in the convex closure of the $\{\mathbf{d}_u\}$ but since each of these $\mathbf{d}_u$ is in the convex closure of R^1, so also is ξ^0, completing the proof of the lemma.

LEMMA 3. *Under the conditions of Theorem 9.3, $R^1 \supset Q^*$.*

The lemma will be proved by defining an expansion of R^1 to the entire space using only points of Q.

By Lemma 2, R^1 contains a cube P with corners $(\pm a, \cdots, \pm a)$. Denote the corners of P by $\{\mathbf{c}_u\}$ with $\mathbf{c}_u$ in the same orthant as $\mathbf{e}_u$. The cube P can be decomposed into disjoint open simplexes of dimension p and less, all vertices being corners of P and every face of every simplex in the collection. A simplex S of dimension q and with vertices $\mathbf{c}_{u(1)}, \cdots, \mathbf{c}_{u(q+1)}$ is defined as

$$S = \left\{\xi : \xi = \sum_{j=1}^{q+1} \theta_{u(j)}\mathbf{c}_{u(j)};\ \sum_{j=1}^{q+1} \theta_{u(j)} = 1, \text{ all } \theta_{u(j)} > 0\right\}.$$

Taking $\theta_u = 0$ for all $u \neq u(j)$ for any j, the $\{\theta_u, u = 0, \cdots, 2^p - 1\}$ are the barycentric coordinates of the point ξ with respect to the simplicial decomposition. The barycentric coordinates are continuous functions over each closed simplex $\bar{S}$ and uniquely defined over P and so are continuous over P. (Cf. Lefschetz [9], p. 97.)

One such decomposition of the cube consists of the $p!$ p-dimensional simplexes $S_i = \{\xi : -a < \xi_{i_1} < \cdots < \xi_{i_p} < a\}$ in which $i = (i_1, \cdots, i_p)$ is any permutation of the integers $(1, \cdots, p)$, and of all faces of the $\{S_i\}$. S_i can be written as

$$S_i = \left\{\xi : \xi = \sum_{j=1}^{p+1} \theta_{u(i,j)}\mathbf{c}_{u(i,j)};\ \sum_{j=1}^{p+1} \theta_{u(i,j)} = 1, \theta_{u(i,j)} > 0\right\}$$

in which $c_{u(i,j)}$ is that corner of P whose i_kth coordinate is $-a$ for $k < j$ and $+a$ for $k \geqq j$. The correspondence between the two representations of S_i is given by:

$$\theta_{u(i,j)} = \frac{1}{2a}(\xi_{i_j} - \xi_{i_{j-1}}), \qquad (j = 1, \cdots, p+1);$$

$$(\xi_{i_0} \equiv -a, \xi_{i_{p+1}} \equiv +a)$$

Any simplex lies entirely in a face, say $\{\xi_i = +a\}$, of P or else does not intersect a face. For if all vertices of S lie in $\{\xi_i = a\}$ then so does S, and if one or more vertices do not lie in $\{\xi_i = a\}$ then the ith coordinate of each point in S is less than $+a$. Let $\mathcal{S}$ be the collection of all simplexes lying in the face F of P. $\mathcal{S}$ is a disjoint simplicial decomposition of F and the barycentric coordinates are continuous over F. Each simplex in $\mathcal{S}$ lies in some single face of the cube. Let S with corners $c_{u(1)}, \cdots, c_{u(k)}$ be an arbitrary member of $\mathcal{S}$ Suppose that S lies in the face $\{\xi_i = a\}$ and hence in M_i^{*+}.

Define a deformation of F as follows For $\xi \varepsilon S$ with $\xi = \sum \theta_{u(j)} c_{u(j)}$; $\sum \theta_{u(j)} = 1$; $\theta_{u(j)} > 0$; define

$$f_t(\xi) = t \sum_{j=1}^{k} \theta_{u(j)} e_{u(j)} + (1 - t) \sum_{j=1}^{k} \theta_{u(j)} c_{u(j)}, \qquad 0 \leqq t \leqq 1$$

$$\left.\begin{aligned} f_t(\xi) &= t \sum_{j=1}^{k} \theta_{u(j)} e_{u(j)} + (1 - t) \sum_{j=1}^{k} \theta_{u(j)} e_{2^p-1-u(j)} \\ &= \sum_{j=1}^{k} \theta_{u(j)} (t e_{u(j)} + (1-t) e_{2^p-1-u(j)}) \end{aligned}\right\} t > 1$$

Let $F(t)$ be the image of F under f_t. The deformation consists in moving each corner of the cube to the corresponding corner of R_i, then out the diagonals.

For each t, f_t is a continuous mapping of F into E_p since the barycentric coordinates are continuous over F. Further, the family of mappings is jointly continuous in t and ξ. f_0 is the identity mapping on F and f_t is homotopic to f_0 for all t. Write $f_t/F \sim f_0/F$; $/F$ to indicate the domain of the mapping and $\sim$ for the homotopy equivalence relation (cf. Lefschetz [9], p. 42).

For $0 \leqq t \leqq 1$, $F(t) \subset R^1$ by Lemma 2 and the convexity of R^1. For $t > 1$, $te_{u(j)} + (1 - t)e_{2^p-1-u(j)}$ lies on the outer diagonal $D(u(j))$. Since the corners of S all lie in M_i^{*+}, so do the $\{e_{u(j)}\}$. Therefore $f_t(S) \subset Q_i^+$, the convex hull of all outer diagonals in M_i^{*+}. Finally then, $F(t) \subset Q$ for all $t > 1$ and

$$P \cup (\bigcup_{t>0} F(t)) \subset R^1 \cup Q.$$

The proof of the lemma will be completed if any point not in P can be shown to lie in $F(t)$ for some $t > 0$.

The distance from the origin to the image $S(t)$ of any simplex S of the face of the cube is never less than a and increases to infinity. For, if S lies in M_i^{*+}, then $c_{u(j),i} = a$, $e_{u(j),i} \geqq a$ and $e_{2^p-1-u(j),i} \leqq -a$. Then for any point ξ in $S(t)$, $\xi_i \geqq a$ if $t \leqq 1$ and $\xi_i \geqq (2t - 1)a$ if $t > 1$.

If x is any point in E_p, denote by π_x the mapping of $E_p - x$ onto the unit sphere centered at the origin, which maps ξ into the projection from the origin of $\xi - x$ (vector subtraction). Denote by π_x/B the mapping π_x with domain restricted to the set B. The following topological theorem is needed:[1]

THEOREM (Hurewicz and Wallman [8], Theorem VI-10). *Let B be a closed bounded subset of E_p. Two points x, y neither contained in B are separated by C, if and only if the mapping π_x/B and π_y/B are not homotopic: $\pi_x/B \not\sim \pi_y/B$.*

Assume there is a point x not contained in $P \cup (\bigcup_{t>0} F(t))$. Since the distance $(0, F(t)) \to \infty$, choose t_1, such that dist. $(0, F(t)) >$ dist. $(0, x)$. Since x does not lie in the cube P, the points 0 and x are separated by the cube boundary F according to the Jordan separation theorem. Applying the topological theorem with $B = F$, $\pi_0/F \not\sim \pi_x/F$. Then for any t

$$\pi_0/F(t) \equiv \pi_0 f_t/F \sim \pi_0 f_0/F \equiv \pi_0/F \not\sim \pi_x/F \equiv \pi_x f_0/F \sim \pi_x f_t/F \equiv \pi_x/F(t),$$

since (a) f_0/F is the identity mapping, (b) $f_0/F \sim f_t/F$ by construction, (c) $\phi f_0/F \sim \phi f_t/F$ by composition for any mapping ϕ with correct domain ([9], p. 42). Since homotopy is an equivalence relation, $\pi_0/F(t) \not\sim \pi_x/F(t)$. Applying the topological theorem again with $B = F(t_1)$, it follows that the points 0 and x are separated by $F(t_1)$. The line segment [0, x] being connected, must intersect $F(t_1)$ which is impossible since dist. $(0, x) <$ dist. $(0, F(t_1))$. Therefore $P \cup (\bigcup_{t>0} F(t)) = E_p$, completing the proof of the lemma and Theorem 9.3.

THEOREM 9.4. *If for each i, $e_i' < 0 < e_i''$ then $R^2 \subset R_I$.*

The line $L_i = \bigcap_{j \neq i} M_j$ intersects F_i^+ at $\xi_i = e_i'' > 0$ and F_i^- at $\xi_i = e_i' < 0$. Approximate F_i^+ and F_i^- by their tangent hyperplanes at these points. If T_i^+ and T_i^- denote the closed half-spaces bounded by these tangent hyperplanes with halves chosen to contain the origin, approximate R_i by $T_i^+ \cap T_i^-$ and R_I by $R^2 = \bigcap_{i=1}^{p} (T_i^- \cap T_i^-)$. Since R_i^{*+} is open and convex and contains the line L_i for $\xi_i > e_i''$, then R_i^{*+} does not intersect the tangent hyperplane or the half-space T_i^+. Similarly, $R_i^{*+} \cap T_i^- = 0$ so that $T_i^+ \cap T_i^- \subset R_i$ and $R^2 \subset R_I$.

10. Acknowledgment. The author is indebted to Professor John W. Tukey for suggesting the subject of this paper, and for his advice and encouragement during its preparation.

REFERENCES

[1] G. E. P. BOX, "The exploration and exploitation of response surfaces: some general considerations and examples," *Biometrics*, Vol. 10 (1954), pp. 16–60.

[2] G. E. P. BOX AND J. S. HUNTER, "A confidence region for the solution of a set of simultaneous equations with an application to experimental design," *Biometrika*, Vol. 41 (1954), pp. 190–198.

[3] G. E. P. BOX AND J. S. HUNTER, "Multi-factor experimental designs for exploring response surfaces," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 195–241.

[4] G. E. P. BOX AND K. B. WILSON, "On the experimental attainment of optimum conditions," *J. Roy. Stat. Soc., Series B*, Vol. 13 (1951), pp. 1–45.

[1] The use of this theorem was indicated to the author by Professor E. H. Spanier.

[5] O. L. DAVIES (Editor), *Design and Analysis of Industrial Experiments*, London, Oliver and Boyd, 1954, Chap. 11.
[6] C. W. DUNNETT AND M. SOBEL, "A bivariate generalization of Student's t-distribution, with tables for certain special cases," *Biometrika*, Vol. 41 (1954), pp. 153–169.
[7] W. FELLER, *Probability Theory and Its Applications*, New York, John Wiley and Sons, 1950.
[8] W. HUREWICZ AND H. WALLMAN, *Dimension Theory*, Princeton, Princeton University Press, 1941.
[9] S. LEFSCHETZ, *Introduction to Topology*, Princeton, Princeton University Press, 1949.
[10] K. PEARSON, *Tables for Statisticians and Biometricians, Part II*, London, Cambridge University Press for the Biometrika Trustees, 1931.
[11] S. N. ROY AND R. C. BOSE, "Simultaneous confidence interval estimation," *Ann. Math. Stat*, Vol. 24 (1953), pp 513–536
[12] H. SCHEFFÉ, "A method for judging all contrasts in the analysis of variance," *Biometrika*, Vol 40 (1953), pp 87–104
[13] J. W. TUKEY, "The problem of multiple comparisons," duplicated manuscript, Princeton University (1953).
[14] S. S. WILKS, *Mathematical Statistics*, Princeton, Princeton University Press, 1943.

EXACT PROBABILITIES AND ASYMPTOTIC RELATIONSHIPS FOR SOME STATISTICS FROM m-th ORDER MARKOV CHAINS[1]

By Leo A. Goodman

University of Chicago

Summary. An exact formula is presented for the probability of a specified frequency count of m-tuples ($m \geqq 1$) in a sequence $X_1, X_2, \cdots, X_N$ from a Markov chain of order $m - 1$ having a denumerable number $a \leqq \infty$ of states. An exact expression is also obtained for the conditional probability of a specified m-tuple count, given the n-tuple count, when the chain is of order $n - 1$ ($n < m$). If $a < \infty$, then this conditional probability, when regarded as a statistic computed from the observed sequence, is shown to be asymptotically equivalent to the product of the probabilities (regarded as a statistic) associated with a corresponding set of a^{n-1} contingency tables with assigned marginals (each table having a^{m-n} row and a columns), where in each table the two attributes described by the table are independent. This fact leads to several simplified tests, related to standard tests of independence in contingency tables, for the null hypothesis H_{n-1} that the Markov chain is of order $n - 1$ within the alternate hypothesis H_{m-1}. Analogous results are also obtained for circular sequences.

1. Introduction. For a circular sequence, Reed Dawson and I. J. Good [4] have presented an exact expression for the conditional probability of a specified frequency count of m-tuples, given the n-tuple count, in the special case where the sequence is stationary and is of so-called zero Markovity; i.e., all $(N - 1)!$ circular permutations of a sequence of N characters are equally likely. It is also proved in [4] that this expression, obtained under the assumption of zero Markovity, is also valid for "negligible" Markovity; i.e., for a stationary chain of order $n - 1$ or less ($n < m$). (The term "Markovity of order m" means that the Markov chain, from which a (linear) sequence of observations is obtained, is of order m; a definition of a "chain of order m" is given in [10] and in Section 3 here. The circular sequence is defined in [4] as a linear stationary sequence with the ends joined.) For a (linear) sequence of N consecutive observations from a stationary chain of order $n - 1$, the conditional probability of a specified m-tuple count, given the n-tuple count, is presented in [4] as the value obtained by augmenting the linear sequence with a blank placed at the end of the sequence, circularizing the augmented sequence, and then applying the exact expression

---

Received May 21, 1957; revised October 18, 1957.

[1] Research carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

I am indebted to I. J. Good and L. J. Savage for some helpful comments, and to R. Dawson and I. J. Good for the opportunity to read [4] before it was published.

for circular sequences to it. The treatment of linear sequences presented in the present paper is more direct, and leads to some different results from those given n [4]. An exact expression is given here for the probability of a specified m-tuple count in a sequence from a chain in the more general case where it need not be stationary and can be of order $m - 1$ (a case of nonnegligible Markovity). An exact formula is also obtained for the conditional probability of a specified m-tuple count $f_{i_1 \cdots i_m}$ in a linear sequence, given the n-tuple counts

$$f_{i_1 \cdots i_n \cdot} = \sum_{i_{n+1}} \sum_{i_{n+2}} \cdots \sum_{i_m} f_{i_1 \cdots i_m} \quad \text{and}$$

$$f_{\cdot i_{m-n+1} \cdots i_m} = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_{m-n}} f_{i_1 \cdots i_m}$$

when the chain need not be stationary and can be of order $n - 1$. Even in the case where the chain is stationary, the formula developed here refers to a different question and is numerically different from that presented in [4]. (In [4], the conditional probability of a specified m-tuple count $f_{i_1 \cdots i_m}$, given the n-tuple count $f_{i_1 \cdots i_n}$, is presented for the stationary chain.) We shall see that, for a (linear) sequence of observations from a chain, it appears to be more relevant to compute the conditional probability when the n-tuple counts $f_{i_1 \cdots i_n \cdot}$ and $f_{\cdot i_{m-n+1} \cdots i_m}$ are given, rather than when the n-tuple count $f_{i_1 \cdots i_n}$ is given.

For stationary circular sequences, it is proved in [4] *that, when the chain is of* zero order and has $a < \infty$ states, then the conditional probability of the observed m-tuple count $\bar{f}_{i_1 \cdots i_m}$, given the 1-tuple count $\bar{f}_i$ (when this probability is regarded as a statistic), is asymptotically equivalent (for N large and $\bar{f}_i/N \to k_i > 0$), to the probability of the cell entries $\bar{f}_{i_1 \cdots i_m}$ in a contingency table with assigned marginals $\bar{f}_{i_1 \cdots i_{m-1}}$ and $\bar{f}_i$, when the two attributes described by the table are independent. In the present paper, this result is generalized to show the asymptotic equivalence, when the chain is of order $n - 1$, between the conditional probability of the observed m-tuple count, given the n-tuple count, and the product of the probabilities of a corresponding set of cell entries in a^{n-1} contingency tables with assigned marginals. An analogous result is also obtained for linear sequences from a chain of order $n - 1$. (The result in [4] for stationary circular sequences of zero order cannot be applied directly to the conditional probability, presented in [4], for the circularized augmented linear sequence (since the 1-tuple count f_B for the augmented blank is 1 and $f_B/N \to 0$); the authors in [4] refer the reader to the present paper for results for linear sequences).

These results lead to the fact that any asymptotic test of contingency for the a^{n-1} independent contingency tables can be used to test the null hypothesis H_{n-1} that the Markov chain is of order $n - 1$ within the alternate hypothesis H_{m-1}. The likelihood ratio test of H_{n-1} within H_n, given by P. G. Hoel [10], can be seen to be of the same form as the joint likelihood ratio test of contingency computed for the a^{n-1} independent contingency tables. The test of H_{n-1} within H_{m-1}, presented by I. J. Good [8] for the circularized sequence, can be seen to be of the same form as the joint likelihood ratio test for the a^{n-1} contin-

gency tables related to the frequency count for the circularized (but not augmented) sequence. (Good also deals with the linear sequence in [8], but he agrees that his paper contains some slips. In applying results obtained for circular sequences to linear sequences, there is a real possibility of errors. (See Corrigenda to [8] and Leo A. Goodman [9].)) For the linear sequence, the likelihood ratio test of H_{n-1} within H_{m-1}, and the χ^2-test of the form used in contingency tables (which is equivalent to the likelihood ratio test), were presented by T. W. Anderson and Leo A. Goodman [1]; but these authors were concerned mainly, in [1], with a large number v of sequences of N consecutive observations from a chain with a finite number of states, where $v \to \infty$ and N was fixed and could, in fact, be small. There was one brief section in [1] dealing with $v = 1$ and $N \to \infty$, and it was based on a long sequence (asymptotic) result, due to M. S. Bartlett [2], concerning the 2-tuple count. The results developed in the present paper are based directly on the exact formula for the distribution of the m-tuple count when $v = 1$ and the chain has denumerably many states.

The approach used here is related to, but different from, earlier work ([1], [2], [6], [13]), where the observed transition proportions were shown to have some properties similar to those of the observed proportions from a set of independent multinomial distributions.

The exact formula developed here for the distribution of the m-tuple count from a chain of order $m - 1$ is a generalization of a result, due to P. Whittle [13], for the special case of $m = 2$. A different, and perhaps simpler, proof of the result in [13] will be presented, and it will be related to the work in [4]. The generalization in the present paper is based directly on this result.

When indicating how many degrees of freedom certain statistics (which were asymptotically χ^2) had, most of the articles mentioned in this section assumed (either explicitly or implicitly) that all the transition probabilities in the Markov chain were positive; for the sake of simplicity, we shall do likewise here when indicating the size of certain contingency tables (and thus how many degrees of freedom the χ^2 statistics corresponding to these tables have). If some of these probabilities are zero, then the methods developed in the present paper can be modified in a straightforward manner to obtain analogous results (see [2]).

2. The 2-tuple and 1-tuple counts. Suppose that a sequence $X_1, X_2, \cdots, X_N$ is obtained from a first order Markov chain with constant transition probability matrix $P = [p_{ij}]$; i.e., the probability is p_{ij} that $X_t = j$, given that $X_{t-1} = i$. For the sake of simplicity, we first assume that the chain has a finite number $a < \infty$ of states. We write f_{ij} for the frequency in the sequence of the 2-tuple (i, j) $(i, j = 1, 2, \cdots, a)$; we also write $\sum_j f_{ij} = f_{i\cdot}$ and $\sum_j f_{ji} = f_{\cdot i}$. If the chain begins in state r and ends in state s ($X_1 = r$ and $X_N = s$), then

$$f_{i\cdot} - f_{\cdot i} = \delta_{ir} - \delta_{is} \qquad (i = 1, 2, \cdots, a), \tag{1}$$

and

$$\sum_i f_{i\cdot} = \sum_i f_{\cdot i} = N - 1, \tag{2}$$

where δ_{ij} equals 1 or 0 according as i and j are equal or unequal. The following result, based on the work in [13], will be used here. Let $T_s(f_{ij})$ be the (sr)th cofactor of the $a \times a$ matrix $[\delta_{ij} - f_{ij}/f_{i\cdot}] = \hat{M}$ if the f_{ij} satisfy (1) and (2), and let it be zero otherwise. (It can be seen that $T_s(f_{ij})$ does not depend on r and is nonnegative.) Then the probability $\prod_{rs}(f_{ij}, s)$ that the 2-tuple count will be $f_{ij}(i, j = 1, 2, \cdots, a)$ and that the sequence ends in s, given that it begins with r, is

$$(3) \qquad T_s(f_{ij}) \frac{\prod_i f_{i\cdot}!}{\prod_i \prod_j f_{ij}!} \prod_i \prod_j p_{ij}^{f_{ij}}.$$

(Actually, it is stated in [13] that (3) is the probability $\prod_{rs}(f_{ij})$ that the 2-tuple count is $f_{ij}(i, j = 1, 2, \cdots, a)$, given that the sequence begins with r and ends with s; this is not quite correct, but can easily be corrected, as has been done here.)

Formula (3) will hold only if $N \geq a$, and $f_{i\cdot} > 0$ and $f_{\cdot i} > 0 (i = 1, 2, \cdots, a)$. However, for $N < a$ or some $f_{i\cdot}$ or $f_{\cdot i}$ equal to 0, (3) still holds if calculated on the basis of a process including only those states that have been observed (see [13]).

A proof of (3), different from that given in [13], will now be presented, since it may increase the understanding of this formula and also since a somewhat different procedure for computing (3) is obtained. This proof uses an approach similar to that applied in [4] to circular sequences with negligible Markovity. It is based on the following combinatorial theorem, called the BEST theorem in [4] (due to N.G. de Bruijn, T. van Aardenne-Ehrenfest, C.A.B. Smith and W.T. Tutte [5]): Given any $a \times a$ matrix $M = [m_{ij}]$ of nonnegative integers, there corresponds an oriented linear graph, with vertices $1, 2, \cdots, a$, such that the number of oriented paths (edges) leading from vertex i to vertex j equals m_{ij}. The matrix, unique to within the same rearrangement of rows as of columns, is called the incidence matrix of the corresponding oriented linear graph. The graph is defined as simple if $m_i = \sum_j m_{ij} = \sum_j m_{ji}$. A circuit in such a graph is defined as a unicursal path passing exactly once through each edge (in the right direction). Let $M' = [m'_{ij}]$ be the $a' \times a'$ matrix formed from M by deleting every row and column consisting wholly of zeros. Then $\sum_j m'_{ij} = \sum_j m'_{ji} = m'_i > 0$, for $i = 1, 2, \cdots, a'$. Let $M^* = [m^*_{ij}]$, where $m^*_{ij} = m'_i\delta_{ij} - m'_{ij}$. Since M^* is a square matrix with each row and column summing to zero, the cofactors of its elements are all equal; let $\| M^* \|$ be the common value of these cofactors. Then the BEST theorem asserts that the number $C(M)$ of distinct circuits, when all the edges are distinguishable, in a simple oriented linear graph with incidence matrix M is

$$C(M) = \| M^* \| \cdot \prod_{i=1}^{a'} (m_i - 1)!.$$

Let $N_{rs}(M)$ be the number of circuits that begin at vertex r and end at vertex s; i.e., the number of paths that pass once through each edge, except for one of the edges leading from vertex s to vertex r. Then, when all the m_{ij} oriented edges

from vertex i to j are distinguishable, we have that $N_{rs}(M) = C(M)m_{sr}$. these edges are not distinguishable, then the number of circuits that begin a vertex r and end at vertex s is $U_{rs}(M) = N_{rs}(M)/\prod_{ij} m_{ij}! = C(M)/\prod_{ij} f_{ij}$ where $f_{sr} = m_{sr} - 1$ and $f_{ij} = m_{ij}$ for $(i, j) \neq (s, r)$; $U_{rs}(M)$ is the number paths that pass directly from vertex i to j in total f_{ij} times, and that begin vertex r and end at s. If $\sum_i \sum_j f_{ij} = N - 1$, the probability of observing an given path (a sequence of vertices or states) that begins at r and ends at s in sequence $X_1, X_2, \cdots X_N$ from a chain with transition probability matr $P = [p_{ij}]$, given that $X_1 = r$, is $\prod_{ij} p_{ij}^{f_{ij}}$. Since the number of such paths $U_{rs}(M)$, the probability of observing one of these paths is

$$\prod_r (f_{ij}, s) = U_{rs}(M) \prod_{ij} p_{ij}^{f_{ij}} = \left[\frac{C(M)}{\prod_{ij} f_{ij}!}\right] \prod_{ij} p_{ij}^{f_{ij}}$$

$$= \left[\frac{T_s(f_{ij}) \prod_{i=1}^{a'} (m_i' - 1)! \prod_{i \neq s} m_i'}{\prod_{ij} f_{ij}!}\right] \prod_{ij} p_{ij}^{f_{ij}}$$

$$= \left[\frac{T_s(f_{ij}) \prod_i f_{i\cdot}!}{\prod_{ij} f_{ij}!}\right] \prod_{ij} p_{ij}^{f_{ij}},$$

where $T_s(f_{ij})$ is the cofactor of an element in the sth row of the matrix M^{**} $[m_{ij}^{**}]$, where $m_{ij}^{**} = m_{ij}^{*}/m_i'$. Q.E.D.

A similar proof was also independently obtained by Dawson and Good in unpublished note. This proof indicates that (3) holds even when a is infini since it depends essentially on the $a' \times a'$ matrix M^*, where a' is finite when is finite, rather than on the $a \times a$ matrix M. Thus, the exact formula, present in [13] for the chain with a finite number of states, also holds for the chain w denumerably many states. (See [6] for some asymptotic distribution theory first order chains with denumerably many states.[2]) This proof also indica that $\prod_r (f_{ij}, s)$ can be computed from the expression $[C(M)/\prod_{ij} f_{ij}!] \prod_{ij} p$ where $m_{ij} = f_{ij}$ for $(i, j) \neq (s, r)$ and $m_{sr} = f_{sr} + 1$, which may sometimes simpler to apply directly than (3).

If r is given, and the f_{ij} satisfy (2) and also (1) for some s, then that s is uniq Thus, s can be determined by (1) as a function of the f_{ij} when r is given. Si the probability $\prod_r (f_{ij})$ of the 2-tuple count f_{ij}, given that $X_1 = r$, is obtai by

$$\prod_r (f_{ij}) = \sum_{x=1}^{a} \prod_r (f_{ij}, x); \tag{4}$$

and since $\prod_r (f_{ij}, x)$ is 0 for all values of $x \neq s$, we have that $\prod_r (f_{ij})$ is eq to (3), if the f_{ij} satisfy (2) and also (1) for some s.

The probability $\prod_r (f_{ij} \mid s)$ of the 2-tuple count $f_{ij}(i, j = 1, 2, \cdots)$, gi that the sequence begins with r and ends with s (which is the verbal (not qu correct) description that was given in [13] for (3)), can actually be obtained

[2] I am indebted to K. L. Chung for bringing [6] to my attention.

dividing $\prod_r(f_{ij}, s)$ by the probability $p_{rs}^{[N-1]}$ that the sequence ends with s: $p_{rs}^{[N-1]}$ is the (rs)th element of the transition probability matrix $P^{N-1} = [p_{ij}^{[N-1]}]$.

If X_1 is a random variable with a probability p_r of being in the rth state, then the probability $\prod(f_{ij}, s, r)$ that the 2-tuple count is f_{ij}, that $X_N = s$ and $X_1 = r$, is simply $\prod_r(f_{ij}, s)p_r$. The general approach given here can also be used to obtain exact expressions for the probability $\prod(f_{ij}, \cdot, r)$ that the 2-tuple count is f_{ij} and that $X_1 = r$, for the probability $\prod(f_{ij}, s)$ that the 2-tuple count is f_{ij} and that $X_N = s$, for the probability $\prod(f_{ij})$ that the 2-tuple count is f_{ij}, and also for various conditional probabilities.

The distribution of the $f_{\cdot j} = \sum_i f_{ij}$ will now be studied, when the chain is of zero order; i.e., $p_{ij} = p_{\cdot j}$ for $i, j = 1, 2, \cdots$. The $f_{\cdot j}$ are the 1-tuple frequencies in the sequence $X_2, X_3, \cdots, X_N$; i.e., $f_{\cdot j}$ is the number of observations in state j among this sequence. The probability $\prod_r(f_{\cdot j}, s)$ that the 1-tuple count in this sequence is $f_{\cdot j}$ and that $X_N = s$, can be derived using the standard multinomial formula, and we obtain

$$(5) \qquad \prod_r(f_{\cdot j}, s) = \left(\frac{f_{\cdot s}}{N-1}\right)\frac{(N-1)!}{\prod_j f_{\cdot j}!}\prod_j p_{\cdot j}^{f_{\cdot j}}.$$

Therefore, for a zero order chain, the conditional probability $\prod_r(f_{ij} \mid f_{\cdot j}, s)$ of the 2-tuple count f_{ij}, given the $f_{\cdot j}$ and s and r, can be obtained by dividing (3) by (5), when the f_{ij} satisfy (1), (2), and also $\sum_i f_{ij} = f_{\cdot j}$. (We can assume, without loss of generality, that $f_{\cdot j}$ and s are such that $\prod_r(f_{\cdot j}, s) > 0$.) Thus

$$(6) \qquad \prod_r(f_{ij} \mid f_{\cdot j}, s) = \left[T_r(f_{ij})\Big/\left(\frac{f_{r\cdot}}{N-1}\right)\right]\left[\frac{\prod_i f_{i\cdot}!\prod_j f_{\cdot j}!}{\prod_i\prod_j f_{ij}!\,(N-1)!}\right]:$$

the second factor is the probability $P(f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ of the cell entries f_{ij} in an ordinary contingency table with assigned marginals $f_{\cdot j}$ and $f_{i\cdot}$, where the two attributes described by the table are independent (see, e.g., [12], p. 278).

Since the $f_{i\cdot}$ can be determined by (1) from the $f_{\cdot j}$, r, and s, we have that (6) is also the conditional probability $\prod_r(f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ of the f_{ij}, given the $f_{\cdot j}$, $f_{i\cdot}$, and r; and (5) is the conditional probability $\prod_r(f_{\cdot j}, f_{i\cdot})$ of the $f_{\cdot j}$ and $f_{i\cdot}$, given r. Furthermore, since the 1-tuple count f_j in the sequence $X_1, X_2, \cdots, X_N$ can be determined from $f_{\cdot j}$ and r by the relation $f_j = f_{\cdot j} + \delta_{jr}$, (6) is also the conditional probability $\prod_r(f_{ij} \mid f_j, s)$ of the 2-tuple count f_{ij}, given the 1-tuple count f_j, and r and s.

From (6), we obtain

$$(7) \qquad T_r(f_{ij})\Big/\left(\frac{f_{r\cdot}}{N-1}\right) = \prod_r(f_{ij} \mid f_{\cdot j}, f_{i\cdot})/P(f_{ij} \mid f_{\cdot j}, f_{i\cdot}).$$

We shall now prove that the statistic (7) converges in probability to unity (that $\prod_r(f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ and $P(f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ are asymptotically equivalent). If the chain is of zero order and $p_{ij} = p_{\cdot j} > 0$. In this case, $f_{r\cdot}/(N-1)$ converges in probability to $p_{r\cdot}$, and it will be necessary to prove only that $T_r(f_{ij})$ converges in probability to $p_{r\cdot}$.

For the sake of simplicity, assume that the chain has a finite number of states and that $f_{i\cdot} > 0$ for $i = 1, 2, \cdots, a$. The $a \times a$ matrix $\hat{M}$ will converge in probability to $M = [\delta_{ij} - p_{\cdot j}]$, and $T_r(f_{ij})$, which is the (sr)th cofactor of $\hat{M}$, will therefore converge in probability to the (sr)th cofactor of M. Since the sum of the entries in each row of M is zero, the cofactors in row s are all equal to the (ss)th cofactor $| M_s |$, the determinant of the $(a - 1) \times (a - 1)$ matrix M_s obtained by deleting the sth row and column in M. By some elementary transformations of M_s, or by the identities between the cofactors and the elements of a matrix (see, e.g., p. 109 in [3]) and the relationship between the principal minors and the characteristic equation (see, e.g., p. 19 in [11]), we see that $| M_s | = p_{\cdot s}$. Hence, $T_r(f_{ij})$ converges in probability to $p_{\cdot s}$. Q.E.D.

This result, concerning the asymptotic equivalence (under the assumption H_0 of zero Markovity) of $\prod_r (f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ and $P(f_{ij} \mid f_{\cdot j}, f_{i\cdot})$, implies that the null hypothesis H_0 can be tested, within H_1, by any asymptotic test of contingency in the contingency table with cell entries f_{ij} and with assigned marginals $f_{\cdot j}$ and $f_{i\cdot}$. This implication follows from an application of the following lemma proved in [4]: If (a) an experiment (with parameter N) has, for each value of N (positive integers tending to infinity) a finite set $F^N = \{F_i^N\}$ of possible outcomes, (b) P_N (or simply P) and P'_N (or simply P') are two probability measures over F^N such that $P'(F_i^N)/P(F_i^N)$ converges in the probability P to unity, where $P'(F_i^N)/P(F_i^N)$ is regarded as a statistic whose distribution is determined by P, and (c) $S(F_i^N)$ is a statistic whose cumulative distribution function Φ_N converges, as N becomes infinite, to a limiting distribution Φ under P, then the distribution function Φ'_N of $S(F_i^N)$ under P' also converges to the same limiting distribution Φ. This lemma can be applied, in order to obtain the desired implication, by taking $P(F_i^N) = [P(f_{ij} \mid f_{\cdot j}, f_{i\cdot}) \prod_r (f_{\cdot j}, f_{i\cdot})]$, and $P'(F_i^N) = \prod_r (f_{ij}, s)$. Since H_0 is assumed, $P'(F_i^N)/P(F_i^N) = \prod_r(f_{ij} \mid f_{\cdot j}, f_{i\cdot})/P(f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ will converge in probability to unity. Since any asymptotic test of contingency in the contingency table with cell entries f_{ij} and with assigned marginals $f_{\cdot j}$ and $f_{i\cdot}$ will have the same asymptotic distribution under $P(f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ (i.e., in the standard case) as under $P(F_i^N)$ (since the $f_{\cdot j}/N$ and $f_{i\cdot}/N$ converge in probability to $p_{\cdot j}$ and $p_{\cdot i}$ respectively), it follows from the lemma that the same standard asymptotic distribution will also hold under $P'(F_i^N)$ (i.e., when H_0 is true).

Since the sequence obtained from the chain is finite, it will not provide estimates of p_{ij}, for all i, j, if the chain has a denumerable infinity of states (see [6]). Thus, when $a = \infty$, select (independently of the data) a finite subset of, say, b states, and consider all states that are not included in this subset as belonging to a single state; i.e., reduce the original number of states to $b + 1 = a'$ in the modified sequence. The tests of H_0, suggested in this section for the case where $a < \infty$, can be applied to the modified sequence consisting of a' states, and the results presented will hold also for this case, A rejection of H_0 for the modified sequence would imply a rejection of this hypothesis for the original chain consisting of denumerably many states. This general method is applied to some different hypotheses relating to Markov chains on p. 293 in [6].

3. The $n + 1$-tuple and the n-tuple counts. Suppose that a sequence $X_1, X_2, \cdots, X_N$ is obtained from a Markov chain of order $n(n \geq 1)$, where the probability is $p_{\mathfrak{i}j}$ that $X_t = j$, given that $(X_{t-n}, X_{t-n+1}, \cdots, X_{t-1}) = (i_1, i_2, \cdots, i_n) = \mathfrak{i}$. For the sake of simplicity, we assume that there are a states in this chain; i.e., X_t can take its value from among a possible values of j. We define a new sequence of random vectors $Z_1 = (X_1, X_2, \cdots, X_n)$, $Z_2 = (X_2, X_3, \cdots, X_{n+1})$, $\cdots$, $Z_{N-n+1} = (X_{N-n+1}, X_{N-n+2}, \cdots, X_N)$ where each vector can take its value from among the a^n possible values of $\mathfrak{i}$. The probability $p_{\mathfrak{i}\mathfrak{f}}$ that $Z_t = \mathfrak{f}$, given that $Z_{t-1} = \mathfrak{i}$, is equal to $p_{\mathfrak{i}j_n}$ for $\mathfrak{i}^* = \mathfrak{j}'$, where $\mathfrak{i}^* = (i_2, i_3, \cdots, i_n)$, $\mathfrak{j}' = (j_1, j_2, \cdots, j_{n-1})$, $\mathfrak{i} = (i_1, \mathfrak{i}^*)$, $\mathfrak{f} = (\mathfrak{j}', j_n)$, and $p_{\mathfrak{i}\mathfrak{f}}$ is zero otherwise. The sequence $Z_1, Z_2, \cdots, Z_{N-n+1}$, is from a first order chain with constant transition probability matrix $P_n = [p_{\mathfrak{i}\mathfrak{f}}]$. This chain has a^n states; P_n is an $a^n \times a^n$ matrix (see [2]).

The frequency $f_{\mathfrak{i}\mathfrak{f}}$ of the 2-tuple $(\mathfrak{i}, \mathfrak{f})$ in the sequence of $(N - n + 1)$ observed Z's gives the $(n + 1)$-tuple frequency $f_{\mathfrak{i}j_n}$ in the sequence of N observed X's for all $(\mathfrak{i}, \mathfrak{f})$ where $\mathfrak{f} = (\mathfrak{i}^*, j_r)$, and $f_{\mathfrak{i}\mathfrak{f}}$ will be zero otherwise. In other words, the frequency $f_{\mathfrak{i}i_{n+1}}$ in the sequence of X's of the $(n + 1)$-tuple $(i_1, i_2, \cdots, i_n, i_{n+1})$ is the number of values of t for which $(X_t, X_{t+1}, X_{t+2}, \cdots, X_{t+n}) = (\mathfrak{i}, i_{n+1})$, *i.e.* the number $f_{\mathfrak{i}\mathfrak{f}}$ of values of t for which $Z_t = (i_1, i_2, \cdots, i_n) = \mathfrak{i}$ and $Z_{t+1} = \mathfrak{f}$, for $\mathfrak{f} = (\mathfrak{i}^*, i_{n+1})$. Since $f_{\mathfrak{i}\mathfrak{f}}$ is the 2-tuple count in a sequence from a first order chain, (3) can be applied to obtain the probability $\prod_{\mathfrak{r}}(f_{\mathfrak{i}\mathfrak{f}}, \mathfrak{s})$ that the $(n + 1)$-tuple count in the observed sequence of X's will be $f_{\mathfrak{i}j_n}$ and that $Z_{N-n+1} = \mathfrak{s}$, given that $Z_1 = \mathfrak{r}$. We obtain

$$(8) \qquad \prod_{\mathfrak{r}}(f_{\mathfrak{i}\mathfrak{f}}, \mathfrak{s}) = T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{f}}) \frac{\prod_{\mathfrak{i}} f_{\mathfrak{i}\cdot}!}{\prod_{\mathfrak{i}}\prod_{\mathfrak{f}} f_{\mathfrak{i}\mathfrak{f}}!} \prod_{\mathfrak{i}}\prod_{\mathfrak{f}} p_{\mathfrak{i}\mathfrak{f}}^{f_{\mathfrak{i}\mathfrak{f}}}$$

$$(9) \qquad = \prod_{\mathfrak{r}}(f_{\mathfrak{i}j}, \mathfrak{s}) = T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{f}}) \frac{\prod_{\mathfrak{i}} f_{\mathfrak{i}\cdot}!}{\prod_{\mathfrak{i}}\prod_{j} f_{\mathfrak{i}j}!} \prod_{\mathfrak{i}}\prod_{j} p_{\mathfrak{i}j}^{f_{\mathfrak{i}j}},$$

where $T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{f}})$ is the $(\mathfrak{s}\mathfrak{r})$th cofactor of the $a^n \times a^n$ matrix $[\delta_{\mathfrak{i}\mathfrak{f}} - f_{\mathfrak{i}\mathfrak{f}}/f_{\mathfrak{i}\cdot}] = \hat{M}_n$. This result could also be obtained by applying the BEST theorem to the vertices $\mathfrak{i}$.

The probability $\prod_{\mathfrak{r}}(f_{\mathfrak{i}\mathfrak{f}}) = \prod_{\mathfrak{r}}(f_{\mathfrak{i}j})$ of the $(n + 1)$-tuple count $f_{\mathfrak{i}j}(j = 1, 2, \cdots, a$, and $\mathfrak{i} = 1, 2, \cdots, a^n)$, given that $Z_1 = \mathfrak{r}$, can be obtained from (9), by applying (1) and (2) to the sequence of Z's. Also, the probability $\prod_{\mathfrak{r}}(f_{\mathfrak{i}\mathfrak{f}} \mid \mathfrak{s}) = \prod_{\mathfrak{r}}(f_{\mathfrak{i}j} \mid \mathfrak{s})$ of the $(n + 1)$-tuple count $f_{\mathfrak{i}j}$, given that $Z_1 = \mathfrak{r}$ and $Z_{N-n+1} = \mathfrak{s}$, can be determined with the aid of (9) and the $(N - n)$th power of P_n.

The distribution of the $f_{\cdot\mathfrak{j}}$ will now be studied, when the sequence of X's is from a chain of order $n - 1(n > 1)$. If the chain is of order $n - 1$ (within the hypothesis H_n), then $p_{\mathfrak{i}j} = p_{\cdot\mathfrak{i}^*j}$ for $i_1, j = 1, 2, \cdots, a$ and for all a^{n-1} values of $\mathfrak{i}^*$.

We define a new sequence of random vectors $W_1 = (X_2, X_3, \cdots, X_n)$, $W_2 = (X_3, X_4, \cdots, X_{n+1})$, $\cdots$, $W_{N-n+1} = (X_{N-n+2}, X_{N-n+3}, \cdots, X_N)$, where each vector can take its value from among the a^{n-1} possible vectors $\mathfrak{i}^*$.

The probability $p_{i^*j^*}$ that $W_t = j^*$, given that $W_{t-1} = i^*$, is equal to $p_{\cdot i^* j_n}$ for $\mathfrak{I} = \hat{j}$, where $\mathfrak{I} = (i_3, i_4, \cdots, i_r)$, $\hat{j} = (j_2, j_3, \cdots, j_{n-1})$, $i^* = (i_2, \mathfrak{I})$, $j^* = (\hat{j}, j_n)$, and $p_{i^*j^*}$ is zero otherwise. We have that $p_{i^*j^*} = p_{\mathfrak{i}\mathfrak{j}}$ for $j_1 = i_2$ and for all values of i_1, where $\mathfrak{i} = (i_1, i^*)$ and $\mathfrak{j} = (j_1, j^*)$. The sequence $W_1, W_2, \cdots, W_{N-n+1}$ is from a first order chain with transition probability matrix $P_{n-1} = [p_{i^*j^*}]$. This chain has a^{n-1} states.

The n-tuple count $f_{\cdot \mathfrak{t}} = g_{\mathfrak{t}}$ in the sequence $X_2, X_3, \cdots, X_N$ can be determined by the 2-tuple count $g_{i^*j^*}$ in the sequence of W's. Also, $g_{i^*j^*} = g_{\mathfrak{t}}$ for $\mathfrak{t} = (j_1, j^*) = (i^*, j_n)$. For the n-tuple count $h_{\mathfrak{t}}$ in the sequence $X_2, X_3, \cdots, X_{N-1}$, we have that $h_{\mathfrak{t}} = g_{\mathfrak{t}} - \delta_{\mathfrak{t}\mathfrak{s}}$. Since $h_{\mathfrak{t}}$ can be determined by the 2-tuple count $h_{i^*j^*}$ in the sequence $W_1, W_2, \cdots, W_{N-n}$ from a first order chain, (3) can be applied to obtain the probability $\prod_{r^*} (h_{i^*j^*}, s')$ that the n-tuple count in the sequence $X_2, X_3, \cdots, X_{N-1}$ will be $h_{\mathfrak{t}}$ and that $W_{N-n} = s'$, given that $W_1 = r^*$. The probability $\prod_{r^*} (g_{\mathfrak{t}}, \mathfrak{s})$ that the n-tuple count in the sequence $X_2, X_3, \cdots, X_N$ will be $g_{\mathfrak{t}}$ and that $Z_{N-n+1} = \mathfrak{s}$, given that $W_1 = r^*$, is simply $\prod_{r^*} (h_{i^*j^*}, s') p_{\cdot s' s_n}$. Thus,

$$\text{(10)} \qquad \prod_{r^*} (g_{\mathfrak{t}}, \mathfrak{s}) = \left[T_{s'}(h_{i^*j^*}) \left(\frac{g_{\mathfrak{s}}}{g_{s'\cdot}} \right) \right] \left[\frac{\prod_{i^*} g_{i^*\cdot}!}{\prod_{\mathfrak{t}} g_{\mathfrak{t}}!} \prod_{i^*} \prod_j p_{\cdot i^* j}^{g_{i^*j}} \right],$$

where $g_{\mathfrak{t}} = h_{\mathfrak{t}} + \delta_{\mathfrak{t}\mathfrak{s}}$.

The $(n + 1)$-tuple count in the sequence $X_1, X_2, \cdots, X_N$ can be denoted by f_{ii^*j} or $f_{ij'j}$, where $i^* = j'$. Also, $\sum_i f_{ii^*j} = f_{\cdot i^* j} = f_{\cdot \mathfrak{t}} = g_{\mathfrak{t}} = g_{i^*j}$ and

$$\sum_j g_{i^*j} = g_{i^*\cdot} = \sum_j f_{\cdot i^* j} = \sum_j \sum_i f_{ii^*j} = f_{\cdot i^* \cdot},$$

where $\mathfrak{i} = (i, i^*)$ and $\mathfrak{t} = (j', j)$. Thus the probability $\prod_{\mathfrak{r}} (f_{\cdot i^* j}, \mathfrak{s})$ that

$$\sum_i f_{ii^*j} = f_{\cdot i^* j}$$

and that $Z_{N-n+1} = \mathfrak{s}$ given that $Z_1 = \mathfrak{r}$, is

$$\text{(11)} \qquad \prod_{\mathfrak{r}} (f_{\cdot i^* j}, \mathfrak{s}) = \left[T_{s'}(h_{i^*j^*}) \left(\frac{f_{\cdot s' s}}{f_{\cdot s' \cdot}} \right) \right] \prod_{i^*} \left[\frac{f_{\cdot i^* \cdot}!}{\prod_j f_{\cdot i^* j}!} \prod_j p_{\cdot i^* j}^{f_{\cdot i^* j}} \right]$$

where $\mathfrak{s} = (s', s)$. Therefore, if the chain is of order $n - 1$, the conditional probability of the $(n + 1)$-tuple count $f_{\mathfrak{i}j} = f_{ii^*j}$, given the $f_{\cdot i^* j}$ and $\mathfrak{s}$ and $\mathfrak{r}$, is obtained by dividing (9) by (11); i.e., $\prod_{\mathfrak{r}} (f_{ii^*j} \mid f_{\cdot i^* j}, \mathfrak{s})$ is equal to

$$\text{(12)} \qquad \left[T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{t}}) / T_{s'}(h_{i^*j^*}) \left(\frac{f_{\cdot s' s}}{f_{\cdot s' \cdot}} \right) \right] \prod_{i^*} \left[\frac{\prod_i f_{ii^*\cdot}! \prod_j f_{\cdot i^* j}!}{\prod_i \prod_j f_{ii^*j}! \, f_{\cdot i^* \cdot}!} \right]$$

(We can assume, without loss of generality, that the $f_{\cdot i^* j}$ and $\mathfrak{s}$ are such that $\prod_{\mathfrak{r}} (f_{\cdot i^* j}, \mathfrak{s}) > 0$.) The second factor in (12) is

$$\prod_{i^*} P_{i^*}(f_{ii^*j} \mid f_{\cdot i^* j}, f_{ii^*\cdot}) = P(f_{ii^*j} \mid f_{\cdot i^* j}, f_{ii^*\cdot}),$$

the product of the probabilities of the cell entries f_{ii^*j} in an $a \times a$ contingency table (for a given $(n - 1)$-tuple i^*), with assigned marginals $f_{\cdot i^* j}$ and $f_{ii^*\cdot}$,

where in each table the two attributes described by the table are independent; i.e., the joint probability of the cell entries $f_{i\mathfrak{r}j}$ for all a^{n-1} independent contingency tables.

It can be seen that (12) is also the conditional probability

$$\prod_{\mathfrak{r}} (f_{i\mathfrak{r}j} \mid f_{\cdot\mathfrak{r}j}, f_{i\mathfrak{r}\cdot})$$

of the $f_{i\mathfrak{r}j}$, given the $f_{\cdot\mathfrak{r}j}$, $f_{i\mathfrak{r}\cdot}$, and $\mathfrak{r}$; it is also the conditional probability $\prod_{\mathfrak{r}} (f_{i\mathfrak{r}j} \mid f_{\mathfrak{r}j}, \mathfrak{s})$ of the $(n+1)$-tuple count $f_{i\mathfrak{r}j}$, given the n-tuple count $f_{\mathfrak{r}j}$, and $\mathfrak{r}$ and $\mathfrak{s}$.

From (12), we have that

$$\left[T_{\mathfrak{s}}(f_{\mathfrak{t}\mathfrak{t}})/T_{s'}(h_{\cdot\mathfrak{r}\cdot}) \left(\frac{f_{\cdot r^{\cdot}}}{f_{\cdot r^{\cdot}}}\right)\right] = \frac{\prod_{\mathfrak{r}}(f_{i\mathfrak{r}j} \mid f_{\cdot\mathfrak{r}j}, f_{i\mathfrak{r}\cdot})}{P(f_{i\mathfrak{r}j} \mid f_{\cdot\mathfrak{r}j}, f_{i\mathfrak{r}\cdot})}. \tag{13}$$

We shall now prove that the statistic (13) converges in probability to unity (thus $T_{\mathfrak{s}}(f_{\mathfrak{t}\mathfrak{t}})$ and $T_{s'}(h_{\cdot\mathfrak{r}\cdot})(f_{\cdot r^{\cdot}}/f_{\cdot r^{\cdot}})$ are asymptotically equivalent, and

$$\prod_{\mathfrak{r}} (f_{i\mathfrak{r}j} \mid f_{\cdot\mathfrak{r}j}, f_{i\mathfrak{r}\cdot})$$

and $P(f_{i\mathfrak{r}j} \mid f_{\cdot\mathfrak{r}j}, f_{i\mathfrak{r}\cdot})$ are also asymptotically equivalent), if the chain is of order $n-1$.

If the chain is of order n (a chain of order $n-1$ is also of order n), we saw earlier that a first order chain could be defined with transition probability matrix $P_n = [p_{\mathfrak{t}\mathfrak{t}}]$, and we shall assume that the asymptotic occupation probabilities $p_{\mathfrak{t}}$ for this first order chain are all positive; i.e., $p_{\mathfrak{t}} > 0$, where $p_{\mathfrak{t}}$ is such that $\sum_{\mathfrak{t}} p_{\mathfrak{t}} p_{\mathfrak{t}\mathfrak{t}} = p_{\mathfrak{t}}$ for all $\mathfrak{t}$. This will be the case if the chain described by P_n is irreducible, (positive) recurrent and aperiodic (see, e.g., [6] and [7]). (If $p_{\mathfrak{t}} = 0$ for some $\mathfrak{t}$, the methods developed in the present paper can be modified in a straightforward manner to obtain analogous results (see [2]).) If the observed sequence is from a chain of order $n-1$, then the occupation probability $p_{\mathfrak{t}} = p_{j'}p_{\cdot j'j}$, where $\mathfrak{t} = (j', j)$, and $p_{j'}$ is the asymptotic occupation probability for the first order chain with transition probability matrix $P_{n-1} = [p_{\cdot\mathfrak{r}j}]$. (Lemma 1 in [6] gives a somewhat different, but related, result for chains with denumerably many states.) Since $\prod_{\mathfrak{r}} (f_{\cdot\mathfrak{r}j}, \mathfrak{s}) > 0$, then $p_{\cdot r^{\cdot}} > 0$ where $\mathfrak{s} = (s', s)$, and $f_{\cdot r^{\cdot}}/f_{\cdot r^{\cdot}}$ will converge in probability to $p_{\cdot r^{\cdot}}$. Thus, it will be necessary to prove only that $T_{\mathfrak{s}}(f_{\mathfrak{t}\mathfrak{t}})/T_{s'}(h_{\cdot j\cdot})$ also converges in probability to $p_{\cdot r^{\cdot}}$.

We have that $T_{\mathfrak{s}}(f_{\mathfrak{t}\mathfrak{t}})$ is the $(\mathfrak{s}\mathfrak{r})$-th cofactor of the matrix $\hat{M}_n$, and $T_{s'}(h_{\cdot\mathfrak{r}\cdot})$ is the $(s'r^*)$th cofactor of the matrix $[\delta_{\cdot j\cdot} - h_{\cdot j\cdot}/h_{\cdot\cdot}] = \hat{M}_{n-1}$. These matrices will converge in probability to the $a^n \times a^n$ matrix $M_n = [\delta_{\mathfrak{t}\mathfrak{t}} - p_{\mathfrak{t}\mathfrak{t}}]$ and the $a^{n-1} \times a^{n-1}$ matrix $M_{n-1} = [\delta_{\cdot j\cdot} - p_{\cdot j\cdot}]$ respectively, and $T_{\mathfrak{s}}(f_{\mathfrak{t}\mathfrak{t}})$ and

$$T_{s'}(h_{\cdot j\cdot})$$

will converge in probability to the $(\mathfrak{s}\mathfrak{r})$th cofactor and the $(s'r^*)$th cofactor of M_n and M_{n-1} respectively. Since in each matrix the sum of the entries in each row is zero, all the cofactors in row $\mathfrak{s}$ of M_n are all equal to the $(\mathfrak{s}\mathfrak{s})$th cofactor $|M_{n,\mathfrak{s}}|$ of M_n, and the cofactors in row s' of M_{n-1} are all equal to the $(s's')$th

$_{\cdot 1j}$ is the n-tuple count for the sequence $X_{m-n+1}, X_{m-n+2}, \cdots, X_N$. Let

$$\Pi_{\mathfrak{r}}(f_{\cdot 1j} \mid f_{\cdot 1j}, \mathfrak{s})$$

be the probability that the m-tuple count will be $f_{\cdot 1j}$, given that $\sum_{\iota} f_{\iota 1j} = f_{\cdot 1j}$ and $(X_1, X_2, \cdots, X_{m-1}) = \mathfrak{r}$ and $(X_{N-m+2}, X_{N-m+3}, \cdots, X_N) = \mathfrak{s}$. If $m = n + 1$, the results in Section 3 give the formula for this probability. If $m = n + 2$, then $\Pi_{\mathfrak{r}}(f_{\iota 1j} \mid f_{\cdot 1j}, \mathfrak{s})$ is equal to

$$[\Pi_{\mathfrak{r}}(f_{\iota 1j} \mid f_{\cdot \iota_2 1j}, \mathfrak{s}) \Pi_{\mathfrak{r}}(f_{\cdot \iota_2 1j} \mid f_{\cdot 1j}, \mathfrak{s})]:$$

the first factor is the probability that the m-tuple count will be $f_{\iota 1j}$, given that $\sum_{\iota_1} f_{\iota 1j} = f_{\cdot \iota_2 1j}$ and $(X_1, X_2, \cdots, X_{m-1}) = \mathfrak{r}$ and $(X_{N-m+2}, \cdots, X_N) = \mathfrak{s}$; the second factor is the probability that the $(m - 1)$-tuple count in the sequence $X_2, X_3, \cdots, X_N$ will be $f_{\cdot \iota_2 1j}$, given that $\sum_{\iota_2} f_{\cdot \iota_2 1j} = f_{\cdot 1j}$ and

$$(X_1, X_2, \cdots, X_{m-1}) = \mathfrak{r}$$

and $(X_{N-m+2}, \cdots, X_N) = \mathfrak{s}$. If the chain is of order n, the results in Section 3 indicate that the first factor is asymptotically equivalent to

$$\prod_{\iota_2 1} \{\prod_{\iota_1} f_{\iota 1 \cdot}! \prod_{j} f_{\cdot \iota_2 1j}! / \prod_{\iota_1} \prod_{j} f_{\iota 1j}! f_{\cdot \iota_2 1}!\}; \tag{15}$$

if the chain is of order $n - 1$, the second factor is asymptotically equivalent to

$$\prod_{1} \{\prod_{\iota_2} f_{\cdot \iota_2 1}! \prod_{j} f_{\cdot 1j}! / \prod_{\iota_2} \prod_{j} f_{\cdot \iota_2 1j}! f_{\cdot 1}!\}, \tag{16}$$

since it can be shown from the derivation of (12) that $\Pi_{\mathfrak{r}}(f_{\cdot \iota_2 1j} \mid f_{\cdot 1j}, \mathfrak{s})$ is asymptotically equivalent to $\Pi_{\mathfrak{r}^*}(f_{\cdot \iota_2 1j} \mid f_{\cdot 1j}, s^*)$, where $\mathfrak{s} = (s_1, s^*)$ and $\mathfrak{r} = (r_1, r^*)$. Thus, for chains of order $n - 1$, $\Pi_{\mathfrak{r}}(f_{\iota 1j} \mid f_{\cdot 1j}, \mathfrak{s})$ for $m = n + 2$ is asymptotically equivalent to the product of (15) and (16); viz.

$$\prod_{1} \{\prod_{\iota} f_{\iota 1 \cdot}! \prod_{j} f_{\cdot 1j}! / \prod_{\iota j} f_{\iota 1j}! f_{\cdot 1}!\}. \tag{17}$$

In the general case where $m > n$, by repeated application of the preceding results for $m = n + 1$ and $n + 2$, we find that, for chains of order $n - 1$,

$$\Pi_{\mathfrak{r}}(f_{\iota 1j} \mid f_{\cdot 1j}, \mathfrak{s})$$

is asymptotically equivalent to (17), the product of the probabilities

$$P_1(f_{\iota 1j} \mid f_{\cdot 1j}, f_{\iota 1 \cdot})$$

of the cell entries $f_{\iota 1j}$ in a contingency table (for a given $(n - 1)$-tuple 1) consisting of a columns ($j = 1, 2, \cdots, a$) and a^{m-n} rows (the a^{m-n} values of ι), with assigned marginals $f_{\cdot 1j}$ and $f_{\iota 1 \cdot}$, when in each table the attributes described by the table (there are a^{n-1} tables) are independent. This result implies that the null hypothesis H_{n-1} can be tested, within the hypothesis H_{m-1}, by any asymptotic test of contingency in the a^{n-1} ordinary $a \times a^{m-n}$ contingency table with cell entries $f_{\iota 1j}$ and with assigned marginals $f_{\cdot 1j}$ and $f_{\iota 1 \cdot}$. These tests

will have $a^{n-1}(a-1)(a^{m-n}-1) = (a^m - a^n)(a-1)/a$ degrees of freedom (see [1] and [8]).

5. **The circular counts.** It was shown in [4] that, for zero order chains, the probability $\prod(\bar{f}_{i_1\cdots i_m} \mid \bar{f}_{i_1\cdots i_n})$ of a specified m-tuple count $\bar{f}_{i_1\cdots i_m}$ in a circular sequence, given the n-tuple count $\bar{f}_{i_1\cdots i_n}(n < m)$, is

(18) $$C([\bar{f}_{i_1\cdots i_m}]) \prod \bar{f}_{i_1\cdots i_n}!/C([\bar{f}_{i_1\cdots i_n}]) \prod \bar{f}_{i_1\cdots i_m}!$$

if $n > 1$, or

$$C([\bar{f}_{i_1\cdots i_m}]) \prod \bar{f}_i!/(N-1)! \prod \bar{f}_{i_1\cdots i_m}! \qquad \text{if } n = 1,$$

where $F^* = [\bar{f}_{i_1\cdots i_n}]$ is the incidence matrix of the graph (see [4]), and $C(M)$ is defined in Section 2 here; (18) is valid for chains of order $n-1$ or less. In the special case $n = 1$ it was proved in [4] that the statistic (18) is asymptotically equivalent to $\prod \bar{f}_{i_1\cdots i_{n-1}}! \prod \bar{f}_i!/N! \prod \bar{f}_{i_1\cdots i_n}!$, the probability

$$P(\bar{f}_{i_1\cdots i_n} \mid \bar{f}_{i_1\cdots i_{n-1}}, \bar{f}_i)$$

of the cell entries $\bar{f}_{i_1\cdots i_n}$ in a contingency table with assigned marginals $\bar{f}_{i_1\cdots i_{n-1}}$ and $\bar{f}_i$. This result for the special case $n = 1$ will now be generalized to the case $n \geqq 1$.

Let us first consider the case where $m = n + 1$. We can write

(19) $$F_n^* = D(\bar{f}_{i_1\cdots i_n}) \left[\delta_{\mathfrak{i}\mathfrak{i}'} - \frac{\bar{f}_{\mathfrak{i}\mathfrak{i}'}}{\bar{f}_{\mathfrak{i}\cdot}}\right],$$

where $\mathfrak{i} = (i_1, i_2, \cdots, i_n)$, $\bar{f}_{\mathfrak{i}\mathfrak{i}'}$ is defined for circular sequences in the same way as $f_{\mathfrak{i}\mathfrak{i}'}$ was defined in Section 3, and $D(\bar{f}_{i_1\cdots i_n})$ is the $a^n \times a^n$ diagonal matrix where the entry in row $\mathfrak{i}$ is $\bar{f}_{\mathfrak{i}}$. We shall assume that no row or column consists wholly of zeros. The common value $|F^*_{n,\mathfrak{i}}|$ of the cofactors of F_n^* can be obtained by determining the $(\mathfrak{i}\mathfrak{i})$th cofactor of $D(\bar{f}_{\mathfrak{i}})$, which is $\prod_{\mathfrak{t}\neq\mathfrak{i}} \bar{f}_{\mathfrak{t}}$, and also the $(\mathfrak{i}\mathfrak{i})$th cofactor of $[\delta_{\mathfrak{i}\mathfrak{i}'} - \bar{f}_{\mathfrak{i}\mathfrak{i}'}/\bar{f}_{\mathfrak{i}\cdot}]$, which converges in probability to the $(\mathfrak{i}\mathfrak{i})$th cofactor $|M_{n,\mathfrak{i}}|$ of M_n. From the results in Section 3, for the case where the chain is of order $n-1$, we see that $|M_{n,\mathfrak{i}}| = |M_{n-1,\mathfrak{i}'}|\, p_{\cdot\mathfrak{i}'i}$, where $\mathfrak{i} = (\mathfrak{i}', i)$. Thus $|F^*_{n,\mathfrak{i}}|$ is asymptotically equivalent to $\prod_{\mathfrak{t}\neq\mathfrak{i}} \bar{f}_{\mathfrak{t}}\, |M_{n-1,\mathfrak{i}'}|\, \bar{f}_{\mathfrak{i}'i}/\bar{f}_{\mathfrak{i}'\cdot}$. Also, the $(\mathfrak{i}'\mathfrak{i}')$th cofactor $|F^*_{n-1,\mathfrak{i}'}|$ of F^*_{n-1} is asymptotically equivalent to $\prod_{j^*\neq\mathfrak{i}'} \bar{f}_{j^*}\, |M_{n-1,\mathfrak{i}'}|$. Hence, $|F^*_{n,\mathfrak{i}}|/|F^*_{n-1,\mathfrak{i}'}|$ is asymptotically equivalent to $\prod_{\mathfrak{t}} \bar{f}_{\mathfrak{t}}/\prod_{j^*} \bar{f}_{j^*}$, and $C(F_n^*)/C(F^*_{n-1})$ is asymptotically equivalent to

$$\prod_{\mathfrak{t}} \bar{f}_{\mathfrak{t}}!/\prod_{j^*} \bar{f}_{j^*}!\,.$$

Therefore, if the chain is of order $n-1$, (18) for $m = n + 1$ is asymptotically equivalent to

(20) $$\frac{\prod_{\mathfrak{t}} \bar{f}_{\mathfrak{t}}!\, \prod_{\mathfrak{t}} \bar{f}_{\mathfrak{t}}!}{\prod_{j^*} \bar{f}_{j^*}!\, \prod_{\mathfrak{t}} \prod_j \bar{f}_{\mathfrak{t}j}!} = \prod_{j^*} \left[\frac{\prod_{j_1} \bar{f}_{j_1 j^*}!\, \prod_j \bar{f}_{j^* j}!}{\prod_{j_1} \prod_j \bar{f}_{j_1 j^* j}!\, \bar{f}_{j^*}!}\right],$$

the product of the probabilities $P_{j^*}(\bar{f}_{j_1 j^* j} \mid \bar{f}_{j_1 j^*}, \bar{f}_{j^* j})$ of the observed cell entries $\bar{f}_{j_1 j^* j}$ in an ordinary $a \times a$ contingency table (for a given $(n-1)$-tuple

j^*), with assigned marginals $\bar{f}_{h_1 j \cdot}$ and $\bar{f}_{j \cdot j}$ (we have that $\sum_j \bar{f}_{h_1 j \cdot j} = \bar{f}_{h_1 j \cdot} = \sum_j \bar{f}_{j h_1 j \cdot} = \bar{f}_{\cdot h_1 j \cdot} = \bar{f}_{h_1 j \cdot}$), where in each table the two attributes described by the table are independent.

This result, concerning asymptotic equivalence in the special case $m = n + 1$, can be applied repeatedly to obtain a general result for the case $m > n$, as was done in Section 4. Thus, if the chain is of order $n - 1$, the statistic

$$\overline{\prod} (\bar{f}_{i_1 \cdots i_m} \mid \bar{f}_{i_1 \cdots i_n})$$

is asymptotically equivalent to

$$\frac{\prod \bar{f}_{i_1 \cdots i_{m-1}}! \prod \bar{f}_{i_1 \cdots i_n}!}{\prod \bar{f}_{i_1 \cdots i_m} \prod \bar{f}_{i_1 \cdots i_{n-1}}!} = \prod_{I} \left[\frac{\prod_\iota \bar{f}_{\iota I}! \prod_i \bar{f}_{Ii}!}{\prod_\iota \prod_i \bar{f}_{\iota I i}! \bar{f}_I!} \right],$$

where $\iota = (i_1, i_2, \cdots, i_{m-n})$, $I = (i_{m-n+1}, i_{m-n+2}, \cdots, i_{m-1})$, $i = i_m$. Hence, any asymptotic test of contingency in the a^{n-1} ordinary $a \times a^{m-n}$ contingency tables (a table for each $(n-1)$-tuple I) with cell entries $\bar{f}_{i_1, \ldots, i_m}$, and with assigned marginals $\bar{f}_{i_1 \cdots i_{m-1}}$ and $f_{i_{m-n+1} \cdots i_m}$ (i.e., $\bar{f}_{i_1 \cdots i_n}$), can be used to test the null hypothesis H_{n-1} within H_{m-1}. The degrees of freedom are as in Section 4.

The reader will note that, in this and the preceding sections, each m-tuple was "split" into an $(m - n)$-tuple ι, an $(n - 1)$-tuple I, and a 1-tuple i; thus obtaining a^{n-1} contingency tables, each $a \times a^{m-n}$. It is possible to split each m-tuple into an $(m - n - r)$-tuple, an $(n - 1)$-tuple, and a $(1 + r)$-tuple $(0 \leqq r \leqq m - n - 1)$; *thus obtaining* a^{n-1} *contingency tables, each* $a^{1+r} \times a^{m-n-r}$. For $r = m - n - 1$, the m-tuple is split into a 1-tuple, an $(n - 1)$-tuple, and a $(m - n)$-tuple; the a^{n-1} contingency tables obtained will differ in general from the a^{n-1} tables obtained for $r = 0$. However, for circular sequences, the product of the likelihood ratios (for testing independence in each table) for the a^{n-1} tables obtained when $r = m - n - 1$ will be equal to the corresponding product for the tables obtained when $r = 0$. For linear sequences, the corresponding products when $r = m - n - 1$ and $r = 0$ will be asymptotically equivalent, under H_{n-1}. Both these products are asymptotically equivalent to the likelihood ratio for testing H_{n-1} within H_{m-1}. Similar remarks could be made about other statistics (e.g., the χ^2 statistic) used to test independence in each table. If the a^{n-1} separate tables were *of interest, the choice between* $r = 0$ or $m - n - 1$ would depend on the alternate hypotheses within H_{m-1} that were in mind.

For $0 \leqq r \leqq m - n - 1$, it can be shown that the asymptotic mean value of the product of the likelihood ratios (when normed in the usual way) is equal to $a^{n-1}(a^{m-n-r} - 1)(a^{1+r} - 1)$, under H_{n-1}. This statistic is not equivalent to the likelihood ratio for testing H_{n-1} within H_{m-1}, unless $r = 0$ or $m - n - 1$. Also, the asymptotic distribution, given H_{n-1}, of this statistic is not χ^2 unless $r = 0$ or $m - n - 1$. For $0 < r < m - n - 1$, the asymptotic distribution, given H_{n-1}, of this statistic is that of a weighted sum (with unequal weights) of χ^2 variates; in this case, the analysis of the a^{n-1} separate contingency tables is not in general as simple and straightforward as when $r = 0$ or $m - n - 1$,

since the usual methods of analysis of contingency tables cannot be applied to this case. This case will be discussed more fully in a later publication by the present author.

6. The exact probability formulas. An illustration will now be presented to indicate the difference between (12), (18), and the formula suggested in [4] for the probability of the specified m-tuple count, given the n-tuple count, in a linear sequence. Consider the special case $a = 2$, $n = 1$, $m = 2$, $N = 5$, and $f_{11} = 0$, $f_{12} = 2$, $f_{21} = 1$, $f_{22} = 1$. Thus $f_{1\cdot} = 2$, $f_{2\cdot} = 2$, $f_{\cdot 1} = 1$, $f_{\cdot 2} = 3$. From (1), we see that $r = 1$, $s = 2$. From (6), the probability of the specified 2-tuple count f_{ij}, given the 1-tuple count $f_{\cdot j}$ and $f_{i\cdot}$ and r, is 2/3. The circularized 2-tuple frequencies are $\bar{f}_{11} = 0$, $\bar{f}_{12} = 2$, $\bar{f}_{21} = 2$, $\bar{f}_{22} = 1$. From (18), the probability of the $\bar{f}_{ij}$, given the $\bar{f}_i$, is 1/2. Using the approach suggested in [4] of applying (18) to the augmented circularized sequence, the probability of the specified 2-tuple count f_{ij}, given the 1-tuple count f_i, is 1/5; this approach yields a correct answer only if the chain is stationary. By listing all possible sequences for $a = 2$ and $N = 5$, the reader will see why different numerical results are obtained for the different probabilities.

REFERENCES

[1] T. W. ANDERSON, AND LEO A. GOODMAN, "Statistical inference about Markov chains," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 89–110.

[2] M. S. BARTLETT, "The frequency goodness of fit test for probability chains," *Proc. Cambridge Philos. Soc.*, Vol. 47, (1951), pp. 86–95.

[3] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[4] R. DAWSON, AND I. J. GOOD, "Exact Markov probabilities from oriented linear graphs," *Ann. Math. Stat.* Vol. 28 (1957), pp. 838–946.

[5] N. G. DE BRUIJN AND T. VAN AARDENNE-EHRENFEST, "Circuits and trees in oriented linear graphs," *Simon Stevin*, Vol. 4 (1950–1), pp. 203–17.

[6] CYRUS DERMAN, "Some asymptotic distribution theory for Markov chains with a denumerable number of states" *Biometrika*, Vol. 43 (1956), pp. 285–94.

[7] WILLIAM FELLER, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, Inc., New York, 1950.

[8] I. J. GOOD, "The likelihood ratio test for Markoff chains," *Biometrika*, Vol. 42 (1955), pp. 531–3; Corrigenda, Vol. 44 (1957), p. 301.

[9] LEO A. GOODMAN, "Simplified runs tests and likelihood ratio tests for Markoff chains," *Biometrika*, Vol. 45 (1958).

[10] P. G. HOEL, "A test for Markoff chains," *Biometrika*, Vol. 41 (1954), pp. 430–3.

[11] C. C. MACDUFFEE, *The Theory of Matrices*, Chelsea Publishing Co., New York, 1946.

[12] A. M. MOOD, *Introduction to the Theory of Statistics*, McGraw-Hill Book Company, New York, 1950.

[13] P. WHITTLE, "Some distribution and moment formulae for the Markov chain," *J. Roy. Stat. Soc., Series B*, Vol. 17 (1955), pp. 235–42.

TESTS OF MULTIPLE INDEPENDENCE AND THE ASSOCIATED CONFIDENCE BOUNDS[1]

By S. N. Roy and R. E. Bargmann

University of North Carolina

1. Summary. In this paper a test based on the union-intersection principle is proposed for overall independence between p variates distributed according to the multivariate normal law, and this is extended to the hypothesis of independence between several groups of variates which have a joint multivariate normal distribution. Methods used in earlier papers [3, 4] have been applied in order to invert these tests for each situation, and to obtain, with a joint confidence coefficient greater than or equal to a preassigned level, simultaneous confidence bounds on certain parametric functions These parametric functions are, in case I, the moduli of the regression vectors: (a) of the variate p on the variates $(p - 1)$, $(p - 2)$, $\cdots$, 2, 1, or on any subset of the latter; (b) of the variate $(p - 1)$ on the variates $(p - 2)$, $(p - 3)$, $\cdots$, 2, 1, or any subset of the latter, etc.; and finally, (c) of the variate 2 on the variate 1. For case II, parallel to each case considered above, there is an analogous statement in which the regression vector is replaced by a regression matrix, β, say, and the "modulus" of the regression vector is replaced by the (positive) square-root of the largest characteristic root of $(\beta\beta')$. Simultaneous confidence bounds on these sets of parameters are given. As far as the proposed tests of hypotheses of multiple independence are concerned they are offered as an alternative to another class of tests based on the likelihood-ratio criterion [5, 6] which has been known for a long time. So far as the confidence bounds are concerned it is believed, however, that no other easily obtainable confidence bounds are available in this area. One of the objects of these confidence bounds is the detection of the "culprit variates" in the case of rejection of the hypothesis of multiple independence, for the "complex" hypothesis is, in this case, the intersection of several more "elementary" hypotheses of two-by-two independence.

2. Introduction, notation, and preliminaries. Case I, which deals with the question of independence among p normally distributed variates, represents a well known situation which has occurred repeatedly in applications. For case II, which deals with the question of independence between k sets of normally distributed variates (where each set contains one or more variates), a number of potential applications has been described by *Wilks* [6]. In addition to the situations mentioned by *Wilks*, an interesting application concerns the problem of "unreliable measurement". If we consider the p_i variates x_i $(i = 1, 2, \cdots, k)$ as

---

Received June 13, 1957.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command.

different measurements of a physically identical quantity which, because of the inaccuracy of the measuring instrument, are not in perfect correlation, the procedures outlined in sections 5 and 6 will study the independence or dependence of the k underlying "true" physical quantities. The "correction-for-attenuation" technique which is widely used in this situation has two serious drawbacks which the present method tries to overcome: (1) It assumes equal error variance for each fallible measurement, and (2) it makes use of a statistic, the "correlation corrected for attenuation", which is not a correlation coefficient, for it may attain values greater than unity. The present method is free from these shortcomings. The confidence bounds discussed in section 6 will then give an indication of the maximum attainable degree of prediction if infallible measurements could be made.

Suppose we have a random sample of size $n + 1$ from an $N[\xi(p \times 1), \Sigma(p \times p)]$, with $p \leqq n$. Then, denoting by $S(p \times p)$ the sample dispersion matrix, we know that S is symmetric and everywhere at least p.s.d., (and also p.d., a.e.). It is also well known that, a.e., there exists a one-to-one transformation from $S(p \times p)$ to $\tilde{T}(p \times p)$ given by $nS = \tilde{T}\tilde{T}'$, where $\tilde{T}$ is a lower triangular matrix with positive diagonal elements. Let t_{ij} $(i \geqq j = 1, 2, \cdots, p)$ denote the elements of $\tilde{T}$, s_{ij} and σ_{ij} $(s_{ij} = s_{ji}, \sigma_{ij} = \sigma_{ji}, i, j = 1, 2, \cdots, p)$ denote the elements of S and Σ, and let s^{ij} and σ^{ij} denote the elements of S^{-1} and Σ^{-1}. Furthermore, let $r_{p\cdot 1,2,\cdots,(p-1)}$, $r_{(p-1)\cdot 1,2,\cdots,(p-2)}$, $\cdots$, $r_{3\cdot 1,2}$, and $r_{2\cdot 1}$ denote, respectively, the multiple correlation coefficient of (p) with $(1, 2, \cdots, p-1)$, of $(p-1)$ with $(1, 2, \cdots, p-2)$, and so on, and finally the simple correlation coefficient of (2) with (1). It may be noted that all except the last are non-negative and a.e. positive. These multiple correlation coefficients will be called the *step-down correlations*. Likewise, let

$$(2.1) \quad \beta'_{i\cdot 1,2,\cdots,i-1}(1 \times \overline{i-1}) = [\sigma_{1i}\sigma_{2i}\cdots\sigma_{i-1,i}]\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1,i-1} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2,i-1} \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{1,i-1} & \sigma_{2,i-1} & \cdots & \sigma_{i-1,i-1} \end{bmatrix}^{-1}$$

(for $i = p, p-1, \cdots, 2$) denote the population regression vector of (i) on $(1, 2, \cdots, i-1)$ and

$$(2.2) \quad \mathbf{b}'_{i\cdot 1,2\cdots i-1}(1 \times \overline{i-1}) = [s_{1i}\, s_{2i}\cdots s_{i-1,i}]\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1,i-1} \\ s_{12} & s_{22} & \cdots & s_{2,i-1} \\ \cdot & \cdot & \cdots & \cdot \\ s_{1,i-1} & s_{2,i-1} & \cdots & s_{i-1,i-1} \end{bmatrix}^{-1}$$

(for $i = p, p-1, \cdots, 2$) denote the corresponding sample regression vector. These regression vectors will be called the *step-down regression vectors*.

Next, we will present the expression for the multiple correlation coefficient by treating it as a special case of a canonical correlation (which will be convenient for later purposes). We have

$$(2.3) \qquad r^2_{i\cdot 1,2,\ldots,i-1} = s_{ii}^{-1}[s_{1i}\cdots s_{i-1,i}]\begin{bmatrix} s_{11} & \cdots & s_{1,i-1} \\ \cdot & \cdots & \cdot \\ s_{1,i-1} & \cdots & s_{i-1,i-1}\end{bmatrix}^{-1}\begin{bmatrix} s_{1i} \\ \cdot \\ s_{i-1,i}\end{bmatrix}$$

for $i = p, p-1, \cdots, 2$. Next we have, by using $nS = \tilde{T}\tilde{T}'$,

$$ns_{ii} = [t_{i1}\cdots t_{ii}][t_{i1}\cdots t_{ii}]' = \sum_{j=1}^{i} t_{ij}^2,$$

$$(2.3.1) \qquad n[s_{1i}\cdots s_{i-1,i}] = [t_{i1}\cdots t_{i,i-1}]\begin{bmatrix} t_{11} & 0 & \cdots & 0 \\ t_{21} & t_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ t_{i-1,1} & \cdot & \cdots & t_{i-1,i-1}\end{bmatrix}'$$

and

$$n\begin{bmatrix} s_{11} & \cdots & s_{1,i-1} \\ \cdot & \cdots & \cdot \\ s_{1,i-1} & \cdots & s_{i-1,i-1}\end{bmatrix} = \begin{bmatrix} t_{11} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ t_{i-1,1} & \cdot & \cdots & t_{i-1,i-1}\end{bmatrix}\begin{bmatrix} t_{11} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ t_{i-1,1} & \cdot & \cdots & t_{i-1,i-1}\end{bmatrix}'$$

It is then easy to check, by substituting the expressions (2.3.1) into (2.3), that

$$(2.4) \qquad r_{i\cdot 1,2,\ldots,i-1} = +\left[\sum_{j=1}^{i-1} t_{ij}^2 \Big/ \sum_{j=1}^{i} t_{ij}^2\right]^{1/2},$$

for $i = p, p-1, \cdots, 2$.

Now, let us turn to the case of a $(p_1 + p_2 + \cdots + p_k)$-variate ($= p$-variate, say) *normal distribution and partition the population dispersion matrix*, $\sum$, into

$$(2.5) \qquad \sum(p \times p) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{12}' & \Sigma_{22} & \cdots & \Sigma_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \Sigma_{1k}' & \Sigma_{2k}' & \cdots & \Sigma_{kk}\end{bmatrix} \begin{matrix} (p_1) \\ (p_2) \\ \\ (p_k)\end{matrix}$$
$$\qquad\qquad (p_1) \quad (p_2) \qquad\quad (p_k)$$

and the sample dispersion matrix, S, into

$$(2.6) \qquad S(p \times p) = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{12}' & S_{22} & \cdots & S_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ S_{1k}' & S_{2k}' & \cdots & S_{kk}\end{bmatrix} \begin{matrix} (p_1) \\ (p_2) \\ \\ (p_k).\end{matrix}$$
$$\qquad\qquad (p_1) \quad (p_2) \qquad\quad (p_k)$$

Regarding each submatrix as an element let us say that there are k "pseudo-rows" and k "pseudo-columns" in the matrices on the right sides of (2.5) and (2.6).

Let $\beta_{i\cdot 1,2,\ldots,i-1}$ and $B_{i\cdot 1,2,\ldots,i-1}$ (for $i = k, k-1, \cdots, 2$) denote the population and the sample regression matrix of the (p_i)-set on the $(p_{i-1} + p_{i-2} + \cdots$

$+ p_1)$-set. These matrices are given by the expressions:

$$\beta_{i\cdot 1,2,\cdots,i-1}(p_i \times \overline{p_{i-1} + p_{i-2} + \cdots + p_1})$$

(2.7)
$$= [\Sigma_{1i}' \Sigma_{2i}' \cdots \Sigma_{i-1,i}'] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots \\ \Sigma_{12}' & \Sigma_{22} & \cdots \\ \cdot & \cdot & \cdots \\ \Sigma_{1,i-1}' & \Sigma_{2,i-1}' & \cdots \end{bmatrix}$$

and

$$B_{i\cdot 1,2,\cdots i-1}(p_i \times \overline{p_{i-1} + p_{i-2} + \cdots + p_1})$$

(2.8)
$$= [S_{1i}' S_{2i}' \cdots S_{i-1,i}'] \begin{bmatrix} S_{11} & S_{12} & \cdots \\ S_{12}' & S_{22} & \cdots \\ \cdot & \cdot & \cdots \\ S_{i,i-1}' & S_{2,i-1}' & \cdots \end{bmatrix}$$

for $i = k, k - 1, \cdots, 2$. The β's and B's will be called the *step-dow matrices*. Also let us denote by

(2.8.1) $$0 \leqq c_{i\cdot 1,2,\cdots,i-1}^{(1)} \leqq c_{i\cdot 1,2,\cdots,i-1}^{(2)} \leqq \cdots \leqq c_{i\cdot 1,2,\cdots,i-1}^{(p_i)} \leqq 1$$

the p_i characteristic roots of the matrix

(2.8.2) $$S_{ii}^{-1}[S_{1i}' S_{2i}' \cdots S_{i-1,i}'] \begin{bmatrix} S_{11} & \cdots & S_{1,i-1} \\ \cdot & \cdots & \cdot \\ S_{1,i-1}' & \cdots & S_{i-1,i-1} \end{bmatrix}^{-1} \begin{bmatrix} S_{1i} \\ \cdot \\ S_{i-1,i} \end{bmatrix},$$

for $i = k, k - 1, \cdots, 2$. It will be noticed that these c's are the sq canonical correlation coefficients of the (p_i)-set with the $(p_1 + p_2 +$ set and that, a.e., the inequalities in (2.8.1) will be strict. For any the p_i canonical correlation coefficients (or rather, as will be see largest of them) will play the same role as $r_{i\cdot 1,2,\cdots,i-1}$ in the previous will be called the *step-down sets of canonical correlations*. We are as for simplicity of discussion, but without loss of generality, that th numbered as to make $\sum_{j=1}^{i-1} p_j \geqq p_i$, (for $i = 2, 3, \cdots, k$). The responding to $\bar{T}$ in the previous case will be introduced in a later se

Sections 3 and 4 will be concerned with the first case, i.e., the variate normal distribution; section 3 will describe a test of the $_0{:}\sigma_{ij} = 0$ $(i \neq j = 1, 2, \cdots, p)$, and section 4 will present simult dence bounds, for $i = p, p - 1, \cdots, 2$, on $(\beta_{i\cdot 1,2,\cdots,i-1}'\beta_{i\cdot 1,2\cdots,i-}$ truncations obtained by deleting any $1, 2, \cdots, (i - 2)$ variates of set).

Sections 5 and 6 will be concerned with the second case, i.e., tha $p_2 + \cdots + p_k)$-variate normal distribution; section 5 will describe hypothesis $H_0{:}\sum_{ij} = 0$ $(i \neq j = 1, 2, \cdots, k)$, and section 6 will pr taneous confidence bounds, for $i = k, k - 1, \cdots, 2$, on the largest c ... $\beta_{i\cdot 1,2,\cdots,i-1}')$ (and on truncations obtained by

$+ p_{i-1}$)-set. These matrices are given by the expressions:

$$\beta_{i\cdot 12\cdots(i-1)}(p_i \times \overline{p_1 + p_2 + \cdots + p_{i-1}})$$

(2.7)
$$= [\Sigma'_{1i}\Sigma'_{2i}\cdots\Sigma'_{i-1,i}]\begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1,i-1} \\ \Sigma'_{12} & \Sigma_{22} & \cdots & \Sigma_{2,i-1} \\ \cdot & \cdot & \cdots & \cdot \\ \Sigma'_{1,i-1} & \Sigma'_{2,i-1} & \cdots & \Sigma_{i-1,i-1}\end{bmatrix}^{-1}$$

and

$$b_{i\cdot 12\cdots(i-1)}(p_i \times \overline{p_1 + p_2 + \cdots + p_{i-1}})$$

(2.8)
$$= [S'_{1i}S'_{2i}\cdots S'_{i-1,i}]\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1,i-1} \\ S'_{12} & S_{22} & \cdots & S_{2,i-1} \\ \cdot & \cdot & \cdots & \cdot \\ S'_{1,i-1} & S'_{2,i-1} & \cdots & S_{i-1,i-1}\end{bmatrix}^{-1},$$

for $i = k, k-1, \cdots, 2$. The β's and b's will be called the *step-down regression matrices*. Next let us denote by

(2.9) $$0 \leqq c^{(1)}_{i\cdot 12\cdots(i-1)} \leqq c^{(2)}_{i\cdot 12\cdots(i-1)} \leqq \cdots \leqq c^{(p_i)}_{i\cdot 12\cdots(i-1)} \leqq 1$$

the p_i characteristic roots of the matrix

(2.10) $$S_{ii}^{-1}[S'_{1i}S'_{2i}\cdots S'_{i-1,i}]\begin{bmatrix} S_{11} & \cdots & S_{1,i-1} \\ \cdot & \cdots & \cdot \\ S'_{1,i-1} & \cdots & S_{i-1,i-1}\end{bmatrix}^{-1}\begin{bmatrix} S_{1i} \\ \cdot \\ S_{i-1,i}\end{bmatrix},$$

for $i = k, k-1, \cdots, 2$. It will be noticed that these c's are the squares of the canonical correlation coefficients of the (p_i)-set with the $(p_1 + p_2 + \cdots + p_{i-1})$-set and that, a.e., the inequalities in (2.9) will be strict. For any (p_i)-set, all the p_i canonical correlation coefficients (or rather, as will be seen later, the largest of them) will play the same role as $r_{i\cdot 12\cdots(i-1)}$ in the previous case. These will be called the *step-down sets of canonical correlations*. We are assuming here for simplicity of discussion, but without loss of generality, that the sets are so numbered as to make $\sum_{j=1}^{i-1} p_j \geqq p_i$ (for $i = 2, 3, \cdots, k$). The matrix corresponding to $\bar{\rho}$ in the previous case will be introduced in a later section.

Sections 3 and 4 will be concerned with the first case, i.e., the case of a p-variate normal distribution; section 3 will describe a test of the hypothesis $\sigma_{ij} = 0$ ($i \neq j = 1, 2, \cdots, p$), and section 4 will present simultaneous confidence bounds, for $i = p, p-1, \cdots, 2$, on $(\rho^2_{i\cdot 12\cdots(i-1)})^{1/2}$ (and on truncations obtained by deleting any $1, 2, \cdots, (i-2)$ variates of the $(i-1)$-set).

Sections 5 and 6 will be concerned with the second case, i.e., that of a $(p_1 + p_2 + \cdots + p_k)$-variate normal distribution; section 5 will describe a test of the hypothesis $\Sigma_{ij} = 0$ ($i \neq j = 1, 2, \cdots, k$), and section 6 will present simultaneous confidence bounds, for $i = k, k-1, \cdots, 2$, on the largest characteristic root of $(\beta_{i\cdot 12\cdots(i-1)}\beta'_{i\cdot 12\cdots(i-1)})$ (and on truncations obtained by deleting any

1, 2, $\cdots$, $(i-2)$ of the sets (p_1), (p_2), $\cdots$, (p_{i-1})). It will be noted that, in case I, the variate (i) is independent of the variates $1, 2, \cdots, i-1$ if and only if $\beta'_{i\cdot 1,2,\cdots,i-1}\beta_{i\cdot 1,2,\cdots,i-1} = 0$, and, in case II, independence of the (p_i)-set on the $[(p_1), (p_2), \cdots, (p_{i-1})]$-set implies, and is implied by, the vanishing of the largest characteristic root of $(\beta_{i\cdot 1,2,\cdots,i-1}\beta'_{i\cdot 1,2,\cdots,i-1})$.

3. Independence in the *p*-variate problem.

3.1. Independence (in distribution) of the step-down correlations, under the null hypothesis. The joint distribution of the t_{ij}'s, for general $\sum$, is well known and given by

$$\text{const}\cdot\exp\left[-\frac{1}{2}\,tr\textstyle\sum^{-1}\tilde{T}\tilde{T}'\right]\prod_{i=1}^{p}t_{ii}^{n-i}\prod_{i\geq j=1}^{p}dt_{ij}. \tag{3.1.1}$$

Among various proofs, a recent one is given in [1, 2]. Under the null hypothesis we have $\sum = D_{\sigma_{ii}}$, where the right side denotes a diagonal matrix with elements $\sigma_{11}, \sigma_{22}, \cdots, \sigma_{pp}$. In this situation, (3.1.1) reduces to

$$\text{const}\cdot\exp\left[-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{i}t_{ij}^{2}\Big/\sigma_{ii}\right]\prod_{i=1}^{p}t_{ii}^{n-i}\prod_{i=1}^{p}\prod_{j=1}^{i}dt_{ij}. \tag{3.1.2}$$

It should be noticed that the t_{ii}'s vary from 0 to ∞, and the t_{ij}'s from $-\infty$ to $+\infty$. Now a comparison with the expression (2.4) shows immediately that the $r_{i\cdot 1,2,\cdots,i-1}$'s $(i = p, \cdots, 2)$ are independently distributed, and their joint distribution is given by

$$\text{const}\cdot\prod_{i=p}^{2}(r^2_{i\cdot 1,2,\cdots,i-1})^{(i-3)/2}(1-r^2_{i\cdot 1,2,\cdots,i-1})^{(n-i-1)/2}\times d(r^2_{i\cdot 1,2,\cdots,i-1}). \tag{3.1.3}$$

3.2. The proposed test and a reason for the allocation of component probabilities. The proposed test is as follows:

$$\begin{aligned}&\text{Accept } H_0 \text{ over } \bigcap_{i=2}^{p}[r^2_{i\cdot 1,2,\cdots,i-1}\leq\mu], \quad\text{and}\\ &\text{reject } H_0 \text{ over } \bigcup_{i=2}^{p}[r^2_{i\cdot 1,2,\cdots,i-1}>\mu],\end{aligned} \tag{3.2.1}$$

where μ is given by

$$\prod_{i=2}^{p}P[r^2_{i\cdot 1,2,\cdots,i-1}\leq\mu\,|\,\rho^2_{i\cdot 1,2,\cdots,i-1}=0] = 1-\alpha. \tag{3.2.2}$$

To obtain μ, proceed as follows: Take a trial value, μ_1, say, between zero and one. Using this value for given n and for $i = 2, 3, \cdots, p$ obtain, from the *Tables of the Incomplete Beta Function*, the probabilities corresponding to the individual factors on the left side of (3.2.2); call these probabilities $\gamma_2, \gamma_3, \cdots, \gamma_p$, say; form the product of the γ_i's and denote it by γ. Proceed in the same manner for other trial values, μ_2, μ_3, etc., and plot the μ_i's against the resulting γ's. Then, on this

plot, μ in (3.2.2) is that value of the μ_i's which corresponds to $\gamma = 1 - \alpha$, the preassigned confidence level.

It is obvious that, if the same μ goes with the different component regions $[r^2_{i\cdot 1,2,\cdots,i-1} \leqq \mu]$, the probability measures that go with these regions are all different. One reason why we make this kind of allocation of the different γ_i's is the following: Notice that the acceptance region for H_0 is the intersection (over $i = p, p-1, \cdots, 2$) of regions $[r^2_{i\cdot 1,2,\cdots,i-1} \leqq \mu]$; but, for a given i, $[r^2_{i\cdot 1,2,\cdots,i-1} \leqq \mu]$ is itself the intersection of regions of the type $[r^2_{i\cdot L(1,2,\cdots,i-1)} \leqq \mu]$, where $r_{i\cdot L(1,2,\cdots,i-1)}$ denotes the (simple) correlation of the ith variate with any linear combination of the variates $(1, 2, \cdots, i-1)$ which includes, as a special case, the variates $(1, 2, \cdots, i-1)$ *individually*. Thus, if we allocate the γ_i's in such a way as to make μ the same for each component region, we attach the same weight not only to the correlations between any pair of the observed variates but also to the correlations between each variate and linear combinations of some others. The reader will perceive that this allocation is not completely symmetric. While symmetry is preserved with respect to all correlations by pairs, the step-down procedure is asymmetric as regards the correlation of any variate with *any* linear combination of all the other variates. However, this is perhaps the best that could be done under this particular approach. It should be noted that, if the square of any simple correlation in the correlation matrix exceeds the value μ, we will have to reject the hypothesis of independence. If, however, the square of the largest correlation coefficient in the correlation matrix stays below μ, we will have to perform the step-down process in order to decide on acceptance or rejection of the hypothesis of independence.

3.3. Relation to the likelihood-ratio test. Since the determinant of R, the correlation matrix, equals $(1 - r^2_{p\cdot 1,2,\cdots,p-1}) \times (1 - r^2_{p-1\cdot 1,2,\cdots,p-2}) \cdots (1 - r^2_{2\cdot 1})$, a test based on the product of the complements of the squares of the step-down correlations is equivalent to the likelihood-ratio test. While the distribution of the determinant of R is fairly complicated, even under the null hypothesis of independence [7], its moments [6] are well known and easily obtained from the joint distribution of the correlation coefficients under the hypothesis of independence. It can be easily verified that they satisfy the recurrence relations:

$$\mu_1' = \prod_{i=1}^{p} \left(1 - \frac{i-1}{n}\right), \tag{3.3.1}$$

and

$$\mu_{\alpha+1}' = \mu_\alpha' \prod_{i=1}^{p} \left(1 - \frac{i-1}{n+2\alpha}\right). \tag{3.3.2}$$

From these relations it is quite simple to obtain the moments, hence the coefficients of skewness and kurtosis, and, from Table 42 of the Biometrika Tables for Statisticians [8], we can obtain, at least for moderately large n, very good approximations to the desired percentage points of the cdf. Thus, for testing the hypothesis of independence, the determinant test is quite useful and closely re-

lated to the step-down procedure presented in this paper. At the moment, however, we do not know of any method to use this determinant test for the construction of confidence bounds on parametric functions without running into complicated non-central distributions, whereas the step-down procedure, as described above, can be *immediately* inverted for the purpose of constructing simultaneous confidence bounds.

4. Confidence bounds associated with the test of independence for a p-variate problem. For shortness, let us now denote by just r_i the $r_{i\cdot 1,2,\cdots,i-1}$ defined by (2.3) and (2.4), by just β_i (*with components* $\beta_{i1}, \beta_{i2}, \cdots, \beta_{i,i-1}$, say) the $\beta_{i\cdot 1,2,\cdots,i-1}$ defined by (2.1), and by just b_i (with components b_{i1}, $b_{i2}, \cdots, b_{i,i-1}$, say) the $b_{i\cdot 1,2,\cdots,i-1}$ defined by (2.2). Assuming a general (symmetric, p.d.) Σ, let us now transform the original variates $x_1, x_2, \cdots, x_p$ to a new set $x_1^*, x_2^*, \cdots, x_p^*$ defined by

$$(4.1)\qquad \begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \\ \cdot \\ x_*^p \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & & 0 \\ -\beta_{21} & 1 & 0 & \cdots & & 0 \\ -\beta_{31} & -\beta_{32} & 1 & \cdots & & 0 \\ \cdot & \cdot & \cdot & \cdots & & \cdot \\ -\beta_{p1} & -\beta_{p2} & -\beta_{p3} & \cdots & -\beta_{p,p-1} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ x_p \end{bmatrix}$$

Then it can be verified by induction or otherwise that the new variates are uncorrelated and hence that the step-down correlations of the new variates, $r_{i\cdot 1,2,\cdots,i-1}^*$, say, $(i = p, p - 1, \cdots, 2)$ are independently distributed. With a joint probability, $1 - \alpha$, say, let us now *make the simultaneous statements:*

$$(4.2)\qquad r_{i\cdot 1,2,\cdots,i-1}^{*2} \leqq \mu \qquad (\text{for } i = p, p - 1, \cdots, 2).$$

We have already seen that, given μ, we can easily find α and also, given α, we can find μ. In order to invert the typical component statement (4.2) and thus make a confidence statement on $\beta_{i\cdot 1,2,\cdots,i-1}$, i.e., on $[\beta_{i1}, \cdots, \beta_{i,i-1}]$, we observe that the multiple correlation coefficient between x_i^* and $[x_1^*, x_2^*, \cdots, x_{i-1}^*]$ *is the same as that between* x_i^* and $[x_1, x_2, \cdots, x_{i-1}]$, *since the starred variates in the* first square brackets are linear combinations of just the non-starred variates the first square brackets are linear combinations of just the non-starred variates in the second square brackets; this fact simplifies our calculation of the desired expressions in terms of the S matrix, i.e., the sample dispersion matrix of the original variates. We may now use the results obtained in reference [4] in connection with the confidence bounds on a (pseudo-) regression matrix of a p-set on a q-set $(p \leqq q)$ for a $(p + q)$-variate *normal distribution*. Let us take expression (3.2) from reference [4] and renumber it as

$$(4.3)\quad c_{\max}^{1/2}(BB') - \lambda c_{\max}^{1/2}(S_{1\cdot 2})c_{\max}^{1/2}(S_{22}^{-1}) \leqq c_{\max}^{1/2}(\beta\beta') \leqq c_{\max}^{1/2}(BB') + \lambda c_{\max}^{1/2}(S_{1\cdot 2})c_{\max}^{1/2}(S_{22}^{-1}),$$

where B and β are the sample and population regression matrices of the p-set on the q-set given, respectively, by $B = S_{12}S_{22}^{-1}$ and $\beta = \sum_{12}\sum_{22}^{-1}$; $S_{1\cdot 2}$ is the sample "residual" matrix of the p-set on the q-set given by $S_{1\cdot 2}$ =

$S_{11} - S_{12}S_{22}^{-1}S_{12}'$; S_{22} is, of course, the sample dispersion matrix of the
Given a preassigned confidence coefficient, $1 - \alpha$, the value of λ in (4.3) c
obtained from the central distribution of the square of the largest can
correlation coefficient, for which recursion formulas are available [2].

For $p = 1$ and $q = i - 1$, (4.3) reduces to

$$(4.4) \quad (\mathbf{b}_i'\mathbf{b}_i)^{1/2} - \lambda(s_{ii} - \mathbf{s}_i'S_{i-1}^{-1}\mathbf{s}_i)^{1/2} c_{max}^{1/2}(S_{i-1}^{-1}) \leqq (\beta_i'\beta_i)^{1/2} \leqq (\mathbf{b}_i'\mathbf{b}_i)^{1/2} + \lambda(s_{ii} - \mathbf{s}_i'S_{i-1}^{-1}\mathbf{s}_i)^{1/2} c_{max}^{1/2}($$

where $\mathbf{b}_i$ and β_i are defined in the opening paragraph of section 4, and

$$\mathbf{s}_i'(1 \times \overline{i-1}) = [s_{1i}, s_{2i}, \cdots, s_{i-1,i}],$$

$$(4.4.1) \quad S_{i-1}(\overline{i-1} \times \overline{i-1}) = \begin{bmatrix} s_{11} & \cdots & s_{1,i-1} \\ \cdot & \cdots & \cdot \\ s_{1,i-1} & \cdots & s_{i-1,i-1} \end{bmatrix},$$

and, finally, $\lambda = +[\mu/(1 - \mu)]^{1/2}$, where μ is obtained by the procedure ou
in the sequel of (3.2.2). It then follows that the typical statement (4.2) $\Rightarrow$
and therefore simultaneous statements (4.2) for $i = p, p - 1, \cdots, 2$ will i
with a joint probability $\geqq 1 - \alpha$, simultaneous confidence bounds (4.
$\beta_i'\beta_i$ for $i = p, p - 1, \cdots, 2$.

In equation (3.1) of reference [4], we may put $p = 1$ and $q = i - 1$ and c
the vector $\mathbf{d}_2$ given there in such a way as to make any one, any two, etc.
finally any $(i - 2)$ components of $\mathbf{d}_2$ equal to zero; if then we make th
responding transition from (3.1) to (3.2) given in reference [4], we will
along with each typical statement (4.4) above, truncated statements wher
one, two, etc., finally any $(i - 2)$ components of β_i and $\mathbf{b}_i$ have been de
without, however, disturbing the expressions that occur with λ. Thus, state
(4.4) and the truncations mentioned above will result in $2^{i-1} - 1$ joint
fidence statements for given i. Since i can take the values $p, p - 1, \cdots,$
will have, altogether, $\sum_{i=2}^{p} (2^{i-1} - 1) = 2^p - p - 1$ confidence statements
a joint confidence coefficient $\geqq 1 - \alpha$.

5. Independence in the $(p_1 + p_2 + \cdots + p_k)$-variate problem.

5.1. Independence (in distribution) of the step-down sets of canonical c lation coefficients, under the null hypothesis. Starting from (2.6) in section
shall make a transformation from S to a partitioned triangular matrix, $\tilde{T}$,
by

$$nS(p \times p) = \begin{matrix} (p_1) \\ (p_2) \\ \\ (p_k) \end{matrix} \begin{bmatrix} \tilde{T}_{11} & 0 & \cdots & 0 \\ T_{21} & \tilde{T}_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{k1} & T_{k2} & \cdots & \tilde{T}_{kk} \end{bmatrix} \begin{bmatrix} \tilde{T}_{11} & 0 & \cdots & 0 \\ T_{21} & \tilde{T}_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{k1} & T_{k2} & \cdots & \tilde{T}_{kk} \end{bmatrix}'$$
$$\qquad\qquad (p_1) \quad (p_2) \qquad (p_k)$$

The distribution of the elements of $\tilde{T}$, under a general $\sum$, will be given by

$$\text{const} \cdot \exp\left[-\frac{1}{2} tr \textstyle\sum^{-1} \tilde{T}\tilde{T}'\right] \prod_{i=1}^{p} t_{ii}^{n-i} \prod_{i \geq j=1}^{k} dT_{ij}, \tag{5.1.1}$$

where $p = p_1 + p_2 + \cdots + p_k$, and $T_{ii} = \tilde{T}_{ii}$ (i.e., triangular). Under $H_0: \sum_{ij} = 0$ $(i \neq j = 1, 2, \cdots, k)$, (5.1.1) reduces to

$$\text{const} \cdot \exp\left[-\frac{1}{2} \sum_{i=1}^{k} tr \textstyle\sum_{ii}^{-1} (T_{i1}, \cdots, \tilde{T}_{ii})(T_{i1}, \cdots, \tilde{T}_{ii})'\right] \prod_{i=1}^{p} t_{ii}^{n-i} \prod_{i \geq j=1}^{k} dT_{ij}. \tag{5.1.2}$$

For shortness, let us write the matrix (2.8.2) in the form

$$S_{ii}^{-1} S^{(i)'} S_{i-1}^{-1} S^{(i)}, \tag{5.1.3}$$

where

$$S^{(i)'}(p_i \times \overline{p_1 + p_2 + \cdots + p_{i-1}}) = \underset{(p_1)}{[S'_{1i}} \quad \underset{(p_2)}{S'_{2i}} \quad \cdots \quad \underset{(p_{i-1})}{S'_{i-1,i}]} \; (p_i), \tag{5.1.4}$$

and

$$S_{i-1}(\overline{p_1 + \cdots + p_{i-1}} \times \overline{p_1 + \cdots + p_{i-1}}) = \begin{bmatrix} S_{11} & \cdots & S_{1,i-1} \\ \cdot & \cdots & \cdot \\ S'_{1,i-1} & \cdots & S_{i-1,i-1} \end{bmatrix}.$$

Also, let us denote the p_i characteristic roots of (5.1.3), ordered from the smallest to the largest, by

$$[c_i^{(1)}, c_i^{(2)}, \cdots, c_i^{(p_i)}].$$

It then follows directly that

$$nS_{ii} = [T_{i1} \cdots \tilde{T}_{ii}][T_{i1} \cdots \tilde{T}_{ii}]', \tag{5.1.5}$$

$$nS^{(i)'} = [T_{i1} \cdots T_{i,i-1}] \begin{bmatrix} \tilde{T}_{11} & 0 & \cdots & 0 \\ T_{21} & \tilde{T}_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{i-1,1} & T_{i-1,2} & \cdots & \tilde{T}_{i-1,i-1} \end{bmatrix}',$$

and

$$nS_{i-1} = \begin{bmatrix} \tilde{T}_{11} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{i-1,1} & \cdot & \cdots & \tilde{T}_{i-1,i-1} \end{bmatrix} \begin{bmatrix} \tilde{T}_{11} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{i-1,1} & \cdot & \cdots & \tilde{T}_{i-1,i-1} \end{bmatrix}'.$$

Substituting the expressions (5.1.5) into (5.1.3) (or (2.8.2)), we see that (5.1.3) reduces to

$$\left[\sum_{j=1}^{i} T_{ij} T'_{ij}\right]^{-1} \left[\sum_{j=1}^{i-1} T_{ij} T'_{ij}\right]. \tag{5.1.6}$$

$S_{11} - S_{12}S_{22}^{-1}S_{12}'$; S_{22} is, of course, the sample dispersion matrix of the q-set. Given a preassigned confidence coefficient, $1 - \alpha$, the value of λ in (4.3) can be obtained from the central distribution of the square of the largest canonical correlation coefficient, for which recursion formulas are available [2].

For $p = 1$ and $q = i - 1$, (4.3) reduces to

$$(4.4)\quad (\mathbf{b}_i'\mathbf{b}_i)^{1/2} - \lambda(s_{ii} - \mathbf{s}_i'S_{i-1}^{-1}\mathbf{s}_i)^{1/2}\, c_{\max}^{1/2}(S_{i-1}^{-1}) \leqq (\beta_i'\beta_i)^{1/2} \leqq (\mathbf{b}_i'\mathbf{b}_i)^{1/2} + \lambda(s_{ii} - \mathbf{s}_i'S_{i-1}^{-1}\mathbf{s}_i)^{1/2} c_{\max}^{1/2}(S_{i-1}^{-1}),$$

where $\mathbf{b}_i$ and β_i are defined in the opening paragraph of section 4, and

$$\mathbf{s}_i'(1 \times \overline{i-1}) = [s_{1i}, s_{2i}, \cdots, s_{i-1,i}],$$

$$(4.4.1)\qquad S_{i-1}\,(\overline{i-1} \times \overline{i-1}) = \begin{bmatrix} s_{11} & \cdots & s_{1,i-1} \\ \cdot & \cdots & \cdot \\ s_{1,i-1} & \cdots & s_{i-1,i-1} \end{bmatrix},$$

and, finally, $\lambda = +[\mu/(1 - \mu)]^{1/2}$, where μ is obtained by the procedure outlined in the sequel of (3.2.2). It then follows that the typical statement (4.2) $\Rightarrow$ (4.4), and therefore simultaneous statements (4.2) for $i = p, p - 1, \cdots, 2$ will imply, with a joint probability $\geqq 1 - \alpha$, simultaneous confidence bounds (4.4) on $\beta_i'\beta_i$ for $i = p, p - 1, \cdots, 2$.

In equation (3.1) of reference [4], we may put $p = 1$ and $q = i - 1$ and choose the vector $\mathbf{d}_2$ given there in such a way as to make any one, any two, etc., and finally any $(i - 2)$ components of $\mathbf{d}_2$ equal to zero; if then we make the corresponding transition from (3.1) to (3.2) given in reference [4], we will have, along with each typical statement (4.4) above, truncated statements where any one, two, etc., finally any $(i - 2)$ components of β_i and $\mathbf{b}_i$ have been deleted without, however, disturbing the expressions that occur with λ. Thus, statement (4.4) and the truncations mentioned above will result in $2^{i-1} - 1$ joint confidence statements for given i. Since i can take the values $p, p - 1, \cdots, 2$, we will have, altogether, $\sum_{i=2}^{p} (2^{i-1} - 1) = 2^p - p - 1$ confidence statements with a joint confidence coefficient $\geqq 1 - \alpha$.

5. Independence in the $(p_1 + p_2 + \cdots + p_k)$-variate problem.

5.1. Independence (in distribution) of the step-down sets of canonical correlation coefficients, under the null hypothesis. Starting from (2.6) in section 2 we shall make a transformation from S to a partitioned triangular matrix, $\widetilde{T}$, given by

$$nS(p \times p) = \begin{matrix} (p_1) \\ (p_2) \\ \\ (p_k) \end{matrix}\begin{bmatrix} \widetilde{T}_{11} & 0 & \cdots & 0 \\ T_{21} & \widetilde{T}_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{k1} & T_{k2} & \cdots & \widetilde{T}_{kk} \end{bmatrix}\begin{bmatrix} \widetilde{T}_{11} & 0 & \cdots & 0 \\ T_{21} & \widetilde{T}_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{k1} & T_{k2} & \cdots & \widetilde{T}_{kk} \end{bmatrix}'$$
$$\qquad (p_1)\quad (p_2)\qquad\quad (p_k)$$

The distribution of the elements of $\tilde{T}$, under a general $\sum$, will be given by

(5.1.1) $$\text{const} \cdot \exp\left[-\frac{1}{2} tr \textstyle\sum^{-1}\tilde{T}\tilde{T}'\right] \prod_{i=1}^{p} t_{ii}^{n-i} \prod_{i \geq j=1}^{k} dT_{ij},$$

where $p = p_1 + p_2 + \cdots + p_k$, and $T_{ii} = \tilde{T}_{ii}$ (i.e., triangular). Under $H_0: \sum_{ij} = 0$ ($i \neq j = 1, 2, \cdots, k$), (5.1.1) reduces to

(5.1.2) $$\text{const} \cdot \exp\left[-\frac{1}{2}\sum_{i=1}^{k} tr \textstyle\sum_{ii}^{-1} (T_{i1}, \cdots, \tilde{T}_{ii})(T_{i1}, \cdots, \tilde{T}_{ii})'\right] \prod_{i=1}^{p} t_{ii}^{n-i} \prod_{i \geq j=1}^{k} dT_{ij}.$$

For shortness, let us write the matrix (2.8.2) in the form

(5.1.3) $$S_{ii}^{-1} S^{(i)\prime} S_{i-1}^{-1} S^{(i)},$$

where

(5.1.4) $$S^{(i)\prime}(p_i \times \overline{p_1 + p_2 + \cdots + p_{i-1}}) = \underset{(p_1)}{[S_{1i}'} \quad \underset{(p_2)}{S_{2i}'} \quad \cdots \quad \underset{(p_{i-1})}{S_{i-1,i}']} \; (p_i),$$

and

$$S_{i-1}(\overline{p_1 + \cdots + p_{i-1}} \times \overline{p_1 + \cdots + p_{i-1}}) = \begin{bmatrix} S_{11} & \cdots & S_{1,i-1} \\ \cdot & \cdots & \cdot \\ S_{1,i-1}' & \cdots & S_{i-1,i-1} \end{bmatrix}.$$

Also, let us denote the p_i characteristic roots of (5.1.3), ordered from the smallest to the largest, by

$$[c_i^{(1)}, c_i^{(2)}, \cdots, c_i^{(p_i)}].$$

It then follows directly that

(5.1.5) $$nS_{ii} = [T_{i1} \cdots \tilde{T}_{ii}][T_{i1} \cdots \tilde{T}_{ii}]',$$

$$nS^{(i)\prime} = [T_{i1} \cdots T_{i,i-1}] \begin{bmatrix} \tilde{T}_{11} & 0 & \cdots & 0 \\ T_{21} & \tilde{T}_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{i-1,1} & T_{i-1,2} & \cdots & \tilde{T}_{i-1,i-1} \end{bmatrix}',$$

and

$$nS_{i-1} = \begin{bmatrix} \tilde{T}_{11} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{i-1,1} & \cdot & \cdots & \tilde{T}_{i-1,i-1} \end{bmatrix} \begin{bmatrix} \tilde{T}_{11} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ T_{i-1,1} & \cdot & \cdots & \tilde{T}_{i-1,i-1} \end{bmatrix}'.$$

Substituting the expressions (5.1.5) into (5.1.3) (or (2.8.2)), we see that (5.1.3) reduces to

(5.1.6) $$\left[\sum_{j=1}^{i} T_{ij}T_{ij}'\right]^{-1}\left[\sum_{j=1}^{i-1} T_{ij}T_{ij}'\right].$$

Now, since by (5.1.2) the sets $[T_{i1} \cdots \tilde{T}_{ii}]$ are independently distributed under the null hypothesis, for $i = k, k-1, \cdots 1$, it follows that the sets of characteristic roots of (5.1.3), viz.,

$$[c_2^{(1)}, \cdots, c_2^{(p_2)}], \quad [c_3^{(1)}, \cdots, c_3^{(p_3)}], \cdots, [c_k^{(1)}, \cdots, c_k^{(p_k)}]$$

are independently distributed.

5.2. *The proposed test of independence.* We propose the following test:

$$\text{(5.2.1)} \qquad \text{Accept } H_0 \text{ over } \bigcap_{i=2}^{k} [c_i^{(p_i)} \leqq \lambda], \quad \text{and}$$

$$\text{reject } H_0 \text{ over } \bigcup_{i=2}^{k} [c_i^{(p_i)} > \lambda],$$

where λ is given by

$$\text{(5.2.2)} \qquad \prod_{i=2}^{k} P[c_i^{(p_i)} \leqq \lambda \mid \gamma_i^{(p_i)} = 0] = 1 - \alpha,$$

and γ denotes the largest characteristic root of $\sum_{ii}^{-1} \sum^{(i)'} \sum_{i-1}^{-1} \sum^{(i)}$, where the $\sum_{ij}$'s are obtained from $\sum$ in exactly the same way as the S_{ij}'s from S in equations (5.1.4). It will be noted that $\gamma_i^{(p_i)}$ is zero if and only if, for given i, $\sum_{ij} = 0$, for $j = 1, 2, \cdots, i-1$. Analogous to section 3.2., we take the same value of λ for each factor on the left-hand side of (5.2.2); the reason is the same as that given in section 3.2. The procedure for obtaining λ is analogous to that given for μ in section (3.2.) except that the incomplete Beta function needs to be replaced by the central distribution function of the (square of) the largest canonical correlation coefficient. The distribution and recursion relations for particular values are discussed explicity in reference [2].

5.3. *Relation to the likelihood-ratio test.* Denoting the jth canonical correlation coefficient of the p_i-set on the $(p_1 + p_2 + \cdots + p_{i-1})$-set by $r^{(j)}_{i\cdot 1,2,\cdots,i-1}$, we see that

$$\text{(5.3.1)} \qquad 1 - r^{(j)2}_{i\cdot 1,2,\cdots,i-1} = c^{(j)}[S_{ii}^{-1}(S_{ii} - S^{(i)'}S_{i-1}^{-1}S^{(i)}]$$
$$= c^{(j)}[R_{ii}^{-1}(R_{ii} - R^{(i)'}R_{i-1}^{-1}R^{(i)}],$$

where $c^{(j)}$ denotes the jth characteristic root, and the R's are the sample correlation matrices corresponding to the covariance matrices, S.
Thus,

$$\text{(5.3.2)} \qquad \prod_{j=1}^{p_i} (1 - r^{(j)2}_{i\cdot 1,2,\cdots,i-1}) = \frac{|R_i|}{|R_{ii}||R_{i-1}|},$$

and the product of the products of all step-down canonical correlation coefficients (or rather, of the complements of their squares) becomes

$$\text{(5.3.3)} \qquad \prod_{i=2}^{k} \prod_{j=1}^{p_i} (1 - r^{(j)2}_{i\cdot 1,2,\cdots,i-1}) = \frac{|R|}{\prod_{i=1}^{k} |R_{ii}|},$$

because $|R_1| = |R_{11}|$. Thus, a comparison with reference [6] shows that a test based on the product of products of all step-down correlations is closely related to the likelihood-ratio test. The distribution of this statistic, under H_0, is discussed in [7], and the moments are given in [6]. They can be readily obtained if, in the joint distribution of all r_{ij}'s (for a general matrix $(\varrho_{ij}) = \hat{R}$, say)

$$P(R \mid \hat{R}) = \frac{p\Gamma(np/2)}{\pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma(n - i + 1/2)} \frac{|R|^{n-p-1/2}}{|\hat{R}|^{n/2}} \tag{5.3.4}$$
$$\times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{d\beta_1 \cdots d\beta_{p-1}}{[tr\hat{R}^{-1}D_\beta R D_\beta]^{np/2}} \prod_{i>j=1}^{p} dr_{ij},$$

(where D_β is a diagonal matrix with elements $e^{-\beta_1/2}$,

$$e^{(\beta_1-\beta_2)/2}, \quad e^{(\beta_2-\beta_3)/2}, \cdots, e^{(\beta_{p-2}-\beta_{p-1})/2}, e^{\beta_{p-1}/2})$$

we set

$$\hat{R} = \begin{bmatrix} \hat{R}_{11} & 0 & \cdots & 0 \\ 0 & \hat{R}_{22} & \cdots & 0 \\ \cdot & & \cdot\cdot & \\ 0 & 0 & & \hat{R}_{kk} \end{bmatrix}.$$

The moments satisfy the convenient recurrence relation:

$$\mu_1' = \frac{\prod_{q=1}^{p} (n - q + 1)}{\prod_{j=1}^{k} \prod_{i=1}^{p_j} (n - i + 1)}, \tag{5.3.5}$$

and

$$\mu_{\alpha+1}' = \mu_\alpha' \frac{\prod_{q=1}^{p} (n + 2\alpha - q + 1)}{\prod_{j=1}^{k} \prod_{i=1}^{p_j} (n + 2\alpha - i + 1)}.$$

Thus, for testing the hypothesis of independence between k sets of normally distributed variates, the determinant test is quite useful and closely related to the step-down procedure. At the moment, however, we cannot easily construct simultaneous confidence bounds on parametric functions on the basis of the determinant test, whereas the step-down procedure can be inverted into a simultaneous confidence statement.

6. Confidence bounds associated with the test of independence for a $(p_1 + p_2 + \cdots + p_k)$-variate problem. Using (5.1.4) let us rewrite $\beta_{i\cdot 1,2,\cdots,i-1}$ in (2.7) and $B_{i\cdot 1,2,\cdots,i-1}$ in (2.8) as

$$\beta_i(p_i \times \overline{p_1 + p_2 + \cdots + p_{i-1}}) = \Sigma^{(i)'} \Sigma_{i-1}^{-1} \tag{6.1}$$

and

$$B_i(p_i \times (p_1 + p_2 + \cdots + p_{i-1})) = S^{(i)'} S_{i-1}^{-1}. \tag{6.2}$$

Next we partition β_i into

$$\begin{matrix} [\beta_{i1} & \beta_{i2} & \cdots & \beta_{i,i-1}] \\ (p_1) & (p_2) & & (p_{i-1}) \end{matrix}$$

and B_i into

$$[B_{i1}\ B_{i2} \cdots B_{i,i-1}].$$

Assuming now a general (symmetric, p.d.) $\sum$, let us transform the original variates $x_1(p_1 \times 1)$, $x_2(p_2 \times 1)$, $\cdots$, $x_k(p_k \times 1)$ into a new set of variates x_1^*, x_2^*, $\cdots$, x_k^* defined by

$$(6.3) \qquad \begin{bmatrix} x_1^* \\ x_2^* \\ \cdot \\ x_k^* \end{bmatrix} \begin{matrix} (p_1) \\ (p_2) \\ \\ (p_k) \end{matrix} = \begin{bmatrix} I(p_1) & 0 & 0 & \cdots & 0 & 0 \\ -\beta_{21} & I(p_2) & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ -\beta_{k1} & -\beta_{k2} & \cdot & \cdots & -\beta_{k,k-1} & I(p_k) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_k \end{bmatrix}$$

Then it can be verified by induction or otherwise that the k sets of new (starred) variates are uncorrelated, and hence the step-down sets of (squares of) canonical correlations, $[c_2^{*(1)}, \cdots, c_2^{*(p_2)}]$, $[c_3^{*(1)}, \cdots, c_3^{*(p_3)}]$, $\cdots$, $[c_k^{*(1)}, \cdots, c_k^{*(p_k)}]$ are independently distributed. With a joint probability, $1 - \alpha$, say, we may thus make the simultaneous statement

$$(6.4) \qquad c_i^{*(p_i)} \leqq \lambda \qquad (\text{for } i = k, k-1, \cdots, 2).$$

Analogous to section 4, with the modifications given in 5.2., we can find λ if α is preassigned. By the same argument as in section 4, and by using (4.3) we can obtain, with a joint confidence coefficient $\geqq 1 - \alpha$, the following sets of simultaneous confidence bounds, for $i = k, k-1, \cdots, 2$:

$$(6.5) \quad c_{\max}^{1/2}(B_iB_i') - \lambda c_{\max}^{1/2}[S_{ii} - S^{(i)\prime}S_{i-1}^{-1}S^{(i)}]c_{\max}^{1/2}(S_{i-1}^{-1}) \leqq c_{\max}^{1/2}(\beta_i\beta_i')$$
$$\leqq c_{\max}^{1/2}(B_iB_i') + \lambda c_{\max}^{1/2}[S_{ii} - S^{(i)\prime}S_{i-1}^{-1}S^{(i)}]c_{\max}^{1/2}(S_{i-1}^{-1}),$$

where β_1 and B_i are defined by (6.1) and (6.2), $S^{(i)}$ and S_{i-1} by (5.1.4), and $\sum^{(i)}$ and $\sum_{i-1}$ analogously. Following the argument presented in section 3 of reference [4] and in section 4 of this paper we see that, with a joint confidence coefficient $\geqq 1 - \alpha$, not only can we make the $(k - 1)$ statements (6.5) but, for each typical statement under (6.5), we can also make a number of truncated confidence statements by deleting any number of variates of the (p_i)-set and any number of variates of the $(p_1 + p_2 + \cdots + p_{i-1})$-set taking care only that the number of variates left in the $(p_1 + \cdots + p_{i-1})$-set is not less than that left in the p_i-set.

REFERENCES

[1] I. OLKIN AND S. N. ROY, "On multivariate distribution theory", *Ann. Math. Stat.*, vol. 25 (1954).
[2] S. N. ROY, "A report on some aspects of multivariate analysis", *Institute of Statistics, University of North Carolina*, mimeograph series No. 121.

[3] S. N. Roy, "Some further results in simultaneous confidence interval estimation", *Ann. Math. Statc.*, vol. 25 (1954).
[4] S. N. Roy and R. Gnanadesikan, "Further contributions to multivariate confidence bounds", *Institute of Statistics, University of North Carolina*, mimeograph series No. 155. Also *Biomitrika*, vol. 44 (1957).
[5] S. S. Wilks, "Certain generalizations in the analysis of variance", *Biometrika*, vol. 24 (1932).
[6] S. S. Wilks, "On the independence of k sets of normally distributed statistical variables", *Econometrica*, vol. 3 (1935)
[7] A. Wald and R. J. Brookner, "On the distribution of Wilks' statistic for testing the independence of several groups of variates", *Ann. Math. Stat.*, vol. 12 (1941).
[8] E. S. Pearson and H. O. Hartley, "Biometrika Tables for Statisticians", vol. I, 4th ed., *Cambridge University Press* (1954)

SEQUENTIAL TESTS FOR VARIANCE RATIOS AND COMPONENTS OF VARIANCE[1]

By Allan Birnbaum

Columbia University, New York, and Imperial College, London

1. Introduction and summary. A general sequential sampling method is given for problems of comparing the variances of two or more normal populations in terms of ratios of variances. Sequential tests are given for a hypothesis specifying the ratio of two variances, including tests for variance components in the analysis of variance (Model II). Such tests provide savings in average required numbers of observations, relative to standard F tests, comparable to those typical of sequential probability ratio tests.

2. Basic sequential sampling rule. Let $x' = (x_1', x_2', \cdots)$, $y' = (y_1', y_2', \cdots)$ be sequences of independent observations from two normal populations with unknown means and unknown respective variances σ_x^2, σ_y^2. Let

$$x_1 = \frac{1}{\sqrt{2}} x_1' - \frac{1}{\sqrt{2}} x_2',$$

$$x_2 = \frac{1}{\sqrt{2 \cdot 3}} x_1' + \frac{1}{\sqrt{2 \cdot 3}} x_2' - \frac{2}{\sqrt{2 \cdot 3}} x_3',$$

$$x_n = \frac{1}{\sqrt{n(n+1)}} x_1' + \cdots + \frac{1}{\sqrt{n(n+1)}} x_n' - \frac{n}{\sqrt{n(n+1)}} x_{n+1}',$$

for $n = 3, 4, \cdots$. Let $y_1, y_2, \cdots$ be similarly defined as functions of $y_1', y_2', \cdots$. Then $x = (x_1, x_2, \cdots)$, $y = (y_1, y_2, \cdots)$ are sequences of independent observations, normally distributed with zero means and respective variances σ_x^2, σ_y^2. (In case the original observations are from populations with known means equal, say, to zero, x and y will denote the sequences of original observations.)

Let g be a given positive number, and let

$$u = (u_1, u_2, \cdots) = (gx_1^2 + gx_2^2, gx_3^2 + gx_4^2, \cdots) \quad \text{and}$$

$$v = (v_1, v_2, \cdots) = (y_1^2 + y_2^2, y_3^2 + y_4^2, \cdots).$$

Let

$$R_i = \sum_{j=1}^{i} u_j \quad \text{and} \quad S_i = \sum_{j=1}^{i} v_j, \qquad \text{for } i = 1, 2, \cdots.$$

Let $T = (T_1, T_2, \cdots)$ be a nondecreasing sequence whose elements are those of $\{R_i\} \cup \{S_i\}$. Since events of the form $R_i = R_j$, $S_i = S_j$, $i \neq j$, and $R_i = S_j$

Received December 14, 1954; revised September, 1956.

[1] Work sponsored under the Office of Naval Research, Contract N6onr-271, T. O. XI, Project 042–034. Reproduction in whole or in part is permitted for any purpose of the United States Government.

have total probability zero, we have with probability one that T is uniquely determined,

$$T_1 = \min(R_1, S_1),$$

$$T_2 = \begin{cases} \min(R_2, S_1) & \text{if } T_1 = R_1, \\ \min(R_1, S_2) & \text{if } T_1 = S_1, \text{etc.} \end{cases}$$

Given T, let $B = (b_1, b_2, \cdots)$ be defined by

$$b_i = \begin{cases} 1 & \text{if } T_i = \text{some } R_j, \\ 0 & \text{if } T_i = \text{some } S_j, \end{cases} \qquad \text{for } i = 1, 2, \cdots.$$

The statistical decision procedures to be described below are based on observed values b_i only.

A simple rule for sequential sampling of x_i's and y_i's so as to obtain the values b_i is the following

Sampling rule 1:

1. Observe u_1 and v_1 (that is, observe x_1, x_2, y_1 and y_2, and compute $u_1 = g(x_1^2 + x_2^2)$ and $v_1 = (y_1^2 + y_2^2)$
2. If an additional observation b_2 is required, then if $u_1 < v_1$, observe u_2, if $u_1 \geqq v_1$, observe v_2
3. Similarly at every stage, if an additional observation b_i is required, then if $\sum u_i < \sum v_i$, observe an additional u_i, while if $\sum u_i \geqq \sum v_i$, observe an additional v_i.
4. Discontinue sampling when the observations $b_1, \cdots, b_m$ thus far obtained suffice to determine a decision according to the particular decision procedure being used

It is clear that to obtain m observations $b_1, \cdots, b_m$, a total of $m + 1$ observations u_i and v_i are required; thus the above rule is most efficient for sampling u_i's and v_i's to obtain b_i's

A minor gain in efficiency here becomes possible if the sampling rule is described in terms of x_i's and y_i's. This follows from the following observation: If only x_1, x_2, and y_1 have been observed, and are such that

$$u_1 \equiv g(x_1^2 + x_2^2) < y_1^2 \leqq v_1,$$

then $u_1 < v_1$ and $b_1 = 1$ are known without need to observe y_2. Similarly if $gx_1^2 > y_1^2 + y_2^2$ is observed, $b_1 = 0$ is known without need to observe x_2. It is clear that the determination of b_1 in this way (that is, by observing whether $u_1 \leqq v_1$, without observing the exact numerical values of both u_1 and v_1) does not alter any mathematical or statistical properties of b_1, since b_1 is defined in terms of an inequality in u_1 and v_1. Analogous observations hold at each stage of sampling, and are the basis for the following

Basic sequential sampling rule:

1. Observe x_1 and y_1.
2. If $gx_1^2 < y_1^2$, observe x_2; if $gx_1^2 \geqq y_1^2$, observe y_2.

TABLE 1

Example of sampling rule (with $g = 1$) and computation of b_i's.

Observed values, in order of sampling		$\Sigma g x_i^2 - \Sigma y_i^2$	b_i values
x_i^2	y_i^2		
1. $x_1^2 = 4$	$y_1^2 = 2$	2	
2.	$y_2^2 = 1$	1	$b_1 = 0$
3.	$y_3^2 = 3$	-2	
4. $x_2^2 = 1$		-1	$b_2 = 1$
5. $x_3^2 = 3$		2	
6.	$y_4^2 = 5$	-3	
7. $x_4^2 = 1$		-2	$b_3 = 1$
8. $x_5^2 = 3$		1	$b_4 = 0$

3. Similarly at every stage, letting $\sum g x_i^2$ and $\sum y_i^2$ denote summations over all observations thus far obtained, if $\sum g x_i^2 < \sum y_i^2$, observe an additional x_i; if $\sum g x_i^2 \geqq \sum y_i^2$, observe an additional y_i.
4. Discontinue sampling when the observations $b_1, \cdots b_m$, thus far obtained suffice to determine a decision according to the particular decision procedure being used.

A convenient tabular method for carrying out the sampling and computing the required b_i's is illustrated in Table 1, for the case $g = 1$ of the sampling rule.

The computation of b_i's may be described thus: As soon as $2r$ or more x_i's have been observed *and* $\sum g x_i^2 - \sum y_i^2 < 0$ is observed, the r^{th} unity value of b_i is observed. As soon as $2r$ or more y_i's have been observed *and* $\sum g x_i^2 - \sum y_i^2 \geqq 0$ is observed, the r^{th} zero value of b_i is observed. In applications where population means are unknown, the relation

$$\sum_1^n x_i^2 = \sum_1^{n+1} \left(x_i' - \frac{1}{n+1} \sum_1^{n+1} x_i' \right)^2 = \sum_1^{n+1} (x_i')^2 - \left(\sum_1^{n+1} x_i' \right)^2 \Big/ (n+1)$$

allows simple application of the method directly to the original observations. It is readily verified that this rule minimizes the number of x_i's and y_i's which must be observed to determine the required b_i's. Since this rule requires at least $(2m + 1)$ and at most $(2m + 2)$ observations x_i and y_i, it affords a saving of at most one such observation (and in terms of expected number of observations, a saving of a fraction of one observation) as compared with the preceding rule.

Hence rule 1 may often be preferred because of its greater simplicity. However only the basic sampling rule will be considered in the following sections.

Clearly the sampling rule depends on the given value of g. Criteria for the choice of g are discussed below.

3. Distribution theory. $\frac{1}{2} u_1, \frac{1}{2} u_2, \cdots$ are independently distributed with common density function

$$f(w) = \frac{1}{g\sigma_x^2} e^{-w/g\sigma_x^2}, \qquad w \geqq 0.$$

Similarly $\frac{1}{2}v_1$, $\frac{1}{2}v_2$, $\cdots$ are independently distributed with common density function

$$h(w) = \frac{1}{\sigma_y^2} e^{-w/\sigma_y^2}, \qquad w \geqq 0.$$

Hence (cf. [1] or [2])

(a) The sequences u and v are distributed as the "waiting times" between successive events in two independent Poisson processes with respective parameters

$$\lambda_1 = \frac{1}{2g\sigma_x^2} \quad \text{and} \quad \lambda_2 = \frac{1}{2\sigma_y^2}.$$

(b) If two such processes are observed simultaneously as one process, the new process is Poisson with mean $\lambda = \lambda_1 + \lambda_2$ and waiting times T_1, $T_2 - T_1$, $T_3 - T_2$, $\cdots$.

(c) $b_i = 1$ (i.e. $T_i =$ some R_j) denotes that the i^{th} event observed in the new process occurred in the first of the two original processes.

(d) The b_i's are independent, with

$$p = \text{Prob}\,\{b_1 = 1\} = \left(1 + \frac{\lambda_2}{\lambda_1}\right)^{-1} = \left(1 + g\,\frac{\sigma_x^2}{\sigma_y^2}\right)^{-1}$$

for all i.

Thus we may apply to B any sequential or nonsequential methods for statistical inferences concerning a binomial parameter p. By use of the relation $\sigma_x^2/\sigma_y^2 = (1/g)\,(1/p - 1)$, any interval estimate of p provides an interval estimate of σ_x^2/σ_y^2, and any procedure for testing a hypothesis on p provides a procedure for testing a hypothesis on σ_x^2/σ_y^2. An unbiased estimate of σ_x^2/σ_y^2 is given by $(1/g)\,(1/\hat{p} - 1)$, where $1/\hat{p}$ is an unbiased estimate of $1/p$ based on inverse binomial sampling of b_i's.

The generalization of the basic sampling rule and distribution theory to the case of comparisons of three or more variances is immediate. In this generalization, the distribution of b_i's would be multinomial instead of binomial.

4. Efficiency. A number of questions concerning the efficiencies of tests based on the above sampling rule are discussed in the following sections.

4a. Comparisons with standard F-tests. The standard method of testing a hypothesis $H_0 : \sigma_x^2/\sigma_y^2 = \rho_0$, for any given $\rho_0 > 0$, is to take fixed numbers n_x, n_y of observations x_i, y_i respectively, and to use the statistic

$$F = \frac{n_y \sum_{i=1}^{n_x} x_i^2}{n_x \rho_0 \sum_{i=1}^{n_y} y_i^2}$$

which under H_0 has the F-distribution with n_x, n_y degrees of freedom. Tables 8.3 and 8.4 of [3] give the operating characteristics of such tests. For example, to test $H_0 : \sigma_x^2/\sigma_y^2 = .3401$ against $H_1 : \sigma_x^2/\sigma_y^2 = (.3401)^{-1} = 2.938$, with $\alpha =$

TABLE 1

Example of sampling rule (with $g = 1$) and computation of b_i's.

Observed values, in order of sampling		$\Sigma g x_i^2 - \Sigma y_i^2$	b_i values
x_i^2	y_i^2		
1. $x_1^2 = 4$	$y_1^2 = 2$	2	
2.	$y_2^2 = 1$	1	$b_1 = 0$
3.	$y_3^2 = 3$	−2	
4. $x_2^2 = 1$		−1	$b_2 = 1$
5. $x_3^2 = 3$		2	
6.	$y_4^2 = 5$	−3	
7. $x_4^2 = 1$		−2	$b_3 = 1$
8. $x_5^2 = 3$		1	$b_4 = 0$

3. Similarly at every stage, letting $\sum g x_i^2$ and $\sum y_i^2$ denote summations over all observations thus far obtained, if $\sum g x_i^2 < \sum y_i^2$, observe an additional x_i; if $\sum g x_i^2 \geqq \sum y_i^2$, observe an additional y_i.
4. Discontinue sampling when the observations $b_1, \cdots b_m$, thus far obtained suffice to determine a decision according to the particular decision procedure being used.

A convenient tabular method for carrying out the sampling and computing the required b_i's is illustrated in Table 1, for the case $g = 1$ of the sampling rule.

The computation of b_i's may be described thus: As soon as $2r$ or more x_i's have been observed *and* $\sum g x_i^2 - \sum y_i^2 < 0$ is observed, the r^{th} unity value of b_i is observed. As soon as $2r$ or more y_i's have been observed *and* $\sum g x_i^2 - \sum y_i^2 \geqq 0$ is observed, the r^{th} zero value of b_i is observed. In applications where population means are unknown, the relation

$$\sum_1^n x_i^2 = \sum_1^{n+1} \left(x_i' - \frac{1}{n+1} \sum_1^{n+1} x_i' \right)^2 = \sum_1^{n+1} (x_i')^2 - \left(\sum_1^{n+1} x_i' \right)^2 \Big/ (n+1)$$

allows simple application of the method directly to the original observations. It is readily verified that this rule minimizes the number of x_i's and y_i's which must be observed to determine the required b_i's. Since this rule requires at least $(2m + 1)$ and at most $(2m + 2)$ observations x_i and y_i, it affords a saving of at most one such observation (and in terms of expected number of observations, a saving of a fraction of one observation) as compared with the preceding rule.

Hence rule 1 may often be preferred because of its greater simplicity. However only the basic sampling rule will be considered in the following sections.

Clearly the sampling rule depends on the given value of g. Criteria for the choice of g are discussed below.

3. Distribution theory. $\frac{1}{2} u_1, \frac{1}{2} u_2, \cdots$ are independently distributed with common density function

$$f(w) = \frac{1}{g\sigma_x^2} e^{-w/g\sigma_x^2}, \qquad w \geqq 0.$$

Similarly $\frac{1}{2}v_1$, $\frac{1}{2}v_2$, $\cdots$ are independently distributed with common density function

$$h(w) = \frac{1}{\sigma_y^2} e^{-w/\sigma_y^2}, \qquad w \geqq 0.$$

Hence (cf. [1] or [2])

(a) The sequences u and v are distributed as the "waiting times" between successive events in two independent Poisson processes with respective parameters

$$\lambda_1 = \frac{1}{2g\sigma_x^2} \quad \text{and} \quad \lambda_2 = \frac{1}{2\sigma_y^2}.$$

(b) If two such processes are observed simultaneously as one process, the new process is Poisson with mean $\lambda = \lambda_1 + \lambda_2$ and waiting times T_1, $T_2 - T_1$, $T_3 - T_2$, $\cdots$.

(c) $b_i = 1$ (i.e. T_i = some R_j) denotes that the i^{th} event observed in the new process occurred in the first of the two original processes.

(d) The b_i's are independent, with

$$p = \text{Prob }\{b_i = 1\} = \left(1 + \frac{\lambda_2}{\lambda_1}\right)^{-1} = \left(1 + g\frac{\sigma_x^2}{\sigma_y^2}\right)^{-1}$$

for all i.

Thus we may apply to B any sequential or nonsequential methods for statistical inferences concerning a binomial parameter p. By use of the relation $\sigma_x^2/\sigma_y^2 = (1/g)(1/p - 1)$, any interval estimate of p provides an interval estimate of σ_x^2/σ_y^2, and any procedure for testing a hypothesis on p provides a procedure for testing a hypothesis on σ_x^2/σ_y^2. An unbiased estimate of σ_x^2/σ_y^2 is given by $(1/g)(1/\hat{p} - 1)$, where $1/\hat{p}$ is an unbiased estimate of $1/p$ based on inverse binomial sampling of b_i's.

The generalization of the basic sampling rule and distribution theory to the case of comparisons of three or more variances is immediate. In this generalization, the distribution of b_i's would be multinomial instead of binomial.

4. Efficiency. A number of questions concerning the efficiencies of tests based on the above sampling rule are discussed in the following sections.

4a. Comparisons with standard F-tests. The standard method of testing a hypothesis $H_0 : \sigma_x^2/\sigma_y^2 = \rho_0$, for any given $\rho_0 > 0$, is to take fixed numbers n_x, n_y of observations x_i, y_i respectively, and to use the statistic

$$F = \frac{n_y \sum_{i=1}^{n_x} x_i^2}{n_x \rho_0 \sum_{i=1}^{n_y} y_i^2}$$

which under H_0 has the F-distribution with n_x, n_y degrees of freedom. Tables 8.3 and 8.4 of [3] give the operating characteristics of such tests. For example, to test $H_0 : \sigma_x^2/\sigma_y^2 = .3401$ against $H_1 : \sigma_x^2/\sigma_y^2 = (.3401)^{-1} = 2.938$, with α =

Type I error = .01 and β = Type II error = .01, a total of $n_x + n_y = 40$ observations suffices provided $n_x = n_y = 20$. If the above sampling rule is used, taking $g = 1$, then H_0 is equivalent to

$$H_0': p = \text{Prob}\,\{b_i = 1\} = p_0 = (1 + .3404)^{-1} = .746$$

and H_1 is equivalent to $H_1': p = p_1 = (1 + 2.938)^{-1} = .254$. By use of binomial probability tables [4] we find Prob $\{\sum_{i=1}^{21} b_i \geqq 11 \mid p = .26\}$ = Prob $\{\sum_{i=1}^{21} b_i \leqq 10 \mid p = .74\} = .0088$; thus a nonsequential test of H_0' against H_1' based on 21 observations b_i has

$$\alpha = \beta < .0088 < .01$$

The sampling rule requires either 43 or 44 observations x_i, y_i to generate 21 observations b_i. Thus the efficiency of the standard F-test is approximately matched by the nonsequential test based on b_i's in this case. (More exact comparisons of efficiency can be made, for example by computing the required exact untabled binomial probabilities and using randomized binomial sample sizes, but this does not seem necessary for present purposes.) Comparisons of this kind are given for a number of other cases in Table 2 below. The properties of the four F-tests are taken from [3], with $n_x = n_y$ in each case. The approximately matching non-sequential binomial tests are each based on the case $g = 1$ of the sampling rule; m is the required binomial sample size in each case. Since for m

TABLE 2

	Standard F-tests					Approximately equivalent non-sequential binomial tests						Value of σ_x^2/σ_y^2 corresponding exactly to H_1'
	Value of σ_x^2/σ_y^2		Strength			Value of p		Strength				
	H_0	H_1	α	β	$n_x + n_y$	H_0'	H_1'	α	β	m	$n_x + n_y \doteq 2m + \frac{3}{2}$	
Test 1	.3404	2.938	.01	.01	40	.73	.27	.0119	.0119	21	43.5	2.704
	.4707	2.124	.05	.05	40	.67	.33	.0520	.0520	21	43.5	2.030
Test 2	1	4.512	.05	.05	40	.5	.15	.0318	.0537	19	39.5	5.667
						.5	.19	.0835	.0532	19	39.5	4.263
						.5	.18	.0577	.0537	20	41.5	4.555
Test 3	1	2.866	.05	.05	80	.5	.26	.0541	.0597	39	79.5	2.846
	1	2.866	.05	.05	80	.5	.25	.0403	.0544	40	81.5	3.000
	1	1.693	.05	.50	80	.5	.37	.0541	.4850	39	79.5	1.703
	1	1.693	.05	.50	80	.5	.36	.0403	.4807	40	81.5	1.778
Test 4	1	4.470	.01	.01	80	.5	.17	.0119	.0099	39	79.5	4.882
		4.470	.01	.01	80	.5	.17	.0083	.0124	40	81.5	4.882
		3.579	.01	.05	80	.5	.21	.0119	.0505	39	79.5	3.762
		3.579	.01	.05	80	.5	.20	.0083	.0432	40	81.5	4.000
		2.114	.01	.50	80	.5	.32	.0019	.4889	39	79.5	2.125
		2.114	.01	.50	80	.5	.31	.0083	.4777	40	81.5	2.226

observations b_i the sampling rule requires $2m + 1$ or $2m + 2$ observations x_i, y_i, $2m + 3/2$ ($\doteq n_x + n_y$) is given for each binomial test for comparison with the $n_x + n_y$ required by the corresponding F-test. In each case investigated, the F-test is approximately matched in efficiency by a nonsequential binomial test based on b_i's.

The F-tests are no doubt preferable to the nonsequential binomial tests, but evidently simplicity of application is the only important basis for this preference. Sequential probability ratio tests [5] are directly applicable to the b_i's. Such tests of $H_0':p = p_0$ against $H_1':p = p_1 > p_0$ require average sample sizes of approximately $m/2$ or less when $p \leq p_0$ and when $p \geq p_1$. Thus by use of the above sampling rule, with application of a sequential test to the b_i's, gains in efficiency over the standard F-test are obtained These gains can be calculated, to close approximation, by use of the average-sample-size function for a sequential binomial test on b_i's corresponding to any given F-test

Two-sided sequential tests on σ_x^2/σ_y^2 based on the b_i's would require two-sided sequential probability ratio tests on a binomial parameter. Such tests are not yet available, but can be constructed by application of the method of sections 4.1.2 and 4.1.3 of [5].

4b. Choice of the scale-factor g. Each of the binomial tests considered in the preceding sections was based on the particular case $g = 1$ of the basic sequential sampling rule. The efficiencies of such binomial tests will in general be still further increased by suitable choices of values of g When a problem of testing a variance ratio $\rho = \sigma_x^2/\sigma_y^2$ is specified by given values of ρ_0, ρ_1, α, and β, it is natural to define a best value of g as follows: Consider the problem of testing $H_0':p = (1 + g\rho_0)^{-1}$ against $H_1':p = (1 + g\,\rho_1)^{-1}$, at strength α, β. Let $n(g)$ be the binomial sample size required for a nonsequential test of strength (at least) α, β. Then a best value of g may be defined as one which minimizes $n(g)$. The calculation of an optimal value of g, by use of binomial probability tables, is elementary.

In the case $\alpha = \beta$, symmetry considerations suggest that $g = (\rho_0\rho_1)^{-1/2}$ is a best value; the same conclusion can be reached more formally by use of the normal approximation to the binomial probability $\alpha = \beta$ This case occurs in some of the examples above: For example, to test $H_0:\rho = .4707$ against $H_1:\rho = (.4707)^{-1} = 2.124$ at strength $\alpha = \beta = .05$, the binomial test based on $g = 1$ requires the (non-sequential) sample size $m = 21$. Taking the non-optimal value $g = 21.24$, the corresponding binomial problem is one of testing $H_0':p = .09$ against $H_1':p = .02$ at strength $\alpha = \beta = .05$, for which a (non-sequential) binomial sample size $m = 50$ is required.

If only ρ_0 and ρ_1 are given, it is interesting to consider whether a value of g exists which is best in the above sense simultaneously for all possible values of (α, β). This is a problem in the "comparison of experiments" [6, pp. 331-6]: For any ρ_0, ρ_1 ($\rho_1 \neq \rho_0$) *and* any two values g_1, g_2 of g ($g_1 \neq g_2$), the "binomial dichotomy" experiment E_1, testing $H_0^{(1)}:p = (1 + g_1\rho_0)^{-1}$ against $H_1^{(1)}:p =$

$(1 + g_1\rho_1)^{-1}$, can be shown to be "not comparable with" the experiment E_2, testing $H_0^{(2)}:p = (1 + g_2\rho_0)^{-1}$ against $H_1^{(2)}:p = (1 + g_2\rho_1)^{-1}$. It follows that a best value of g depends on the desired strength (α, β) as well as on ρ_0, ρ_1 in any particular testing problem.

While the value $g = 1$ may not be optimal in those testing problems of Table 1 in which $\rho_1 \neq \rho_0^{-1}$ or $\alpha \neq \beta$, it is evident that this value suffices to provide the savings in expected numbers of observations pointed out in Section 4a above.

4c. Comparison with other sampling rules. It is of interest to compare the above sampling rule with the one considered by Girshick in [7, pp. 134–6]. Girshick's rule is: observe x_1 and y_1; if more observations are needed, observe x_2 and y_2; continue taking such pairs of observations (x_i, y_i) as long as required to terminate a particular inference procedure. (For non-sequential F tests of the hypothesis $H_0:\sigma_x^2/\sigma_y^2 = \rho_0 < 1$ against $H_1:\sigma_x^2/\sigma_y^2 = \rho_0^{-1}$, with a total of $2n$ observations and equal Type I and Type II error rates, $n_x = n_y = n$ are optimal sample sizes. This case is formally the instance of Girshick's sampling rule in which the observation of exactly n pairs (x_i, y_i) is prescribed.) Girshick gave an optimal sequential probability ratio test based on this sampling rule for deciding which of the two variances σ_x^2, σ_y^2 is the larger, and showed that the power functions of such tests depend just on $(1/\sigma_x^2 - 1/\sigma_y^2)$, a parameter which is not of interest in most applications. This suggests that in order to obtain a test whose power function depends just on the variance ratio (which is generally the parameter of interest), we must

(a) use Girshick's sampling rule and apply a test other than Girshick's, which must then lack certain efficiency properties of the sequential probability ratio test, or

(b) use some other sampling rule as a basis for a test.

Rushton [8] and Johnson [9] have given procedures which represent alternative (a). Johnson's Procedure I is based on Girshick's sampling rule and the sequence of statistics

$$\frac{x_1^2}{y_1^2}, \quad \frac{x_1^2 + x_2^2}{y_1^2 + y_2^2}, \cdots \frac{\sum_1^r x_i^2}{\sum_1^r y_i^2}, \cdots,$$

and consists of a sequential probability ratio test applied to this sequence of non-independent statistics. The average sample sizes required by this procedure are not known. Johnson also gives some alternative procedures based on Girshick's sampling rule which are evidently less efficient but have approximately-known average sample sizes.

Alternative (b) is represented by the sampling rule and tests of the preceding sections.

This comparison of sampling rules illustrates a general problem of designing sequential sampling rules which are appropriate and optimal for various problems of testing composite hypotheses. Other illustrations are provided by various procedures for comparing Poisson processes [2]. Some general methods for the design of appropriate sampling rules will be given in another paper.

The above sampling rule and tests illustrate a remark in [2, p. 257]: "··· problems dealing with variances of normal populations have direct analogues in problems dealing with parameters of Poisson processes. ··· " The above methods are analogues of Methods 1–3, pp. 261–2, in [2]

4d. Near-optimal properties. Consider the problem of testing $H_0: \sigma_x^2/\sigma_y^2 = \rho_0$ against $H_1: \sigma_x^2/\sigma_y^2 = \rho_1$, $\rho_1 > \rho_0 > 0$, at a given significance level α. *Suppose that sequential sampling is conducted according to the above* basic sampling rule, with any given value of $g > 0$, and any given termination rule which with probability one leads to termination. Then, relative to the *given* sampling and termination procedure (i.e relative to the corresponding given sample space), the present problem may be viewed as one of nonsequential testing between two composite hypotheses, on the basis of a single (vector) observation. The parameter space consists of points (ρ, τ), where $\rho = \sigma_x^2/\sigma_y^2$, and $\tau = \sigma_y^2$.

Consider the simple hypothesis $H_0^*: (\rho, \tau) = (\rho_0, \tau_0)$, and the simple alternative $H_1^*: (\rho, \tau) = (\rho_1, \tau_1)$, where $\tau_1 = \tau_0 (\rho_0(\rho_1 + 1))/(\rho_1(\rho_0 + 1))$, and τ_0 has any fixed positive value. By the Neyman-Pearson lemma, every best test of H_0^* against H_1^* has a critical region of the form

$$W_k = \{(x, y) \mid \lambda(x, y) \geq k\},$$

for some given $k \geq 0$, where

$$\lambda(x, y) = \frac{f(x_1, \cdots x_{n_x}, y_1, \cdots y_{n_y} \mid H_1^*)}{f(x_1, \cdots x_{n_x}, y_1, \cdots y_{n_y} \mid H_0^*)}$$

is the ratio of likelihood functions Such a test has significance level α^* which, when k is suitably chosen, equals the prescribed α. This test then has power $1 - \beta^*$ under H_1^* which is the maximum attainable under the stated restrictions.

On the other hand, a best test of H_0^* against H_1^* based only on the observed values of b_i's provided by the given sampling procedure has a rejection region of the form

$$W_k' = \{(x, y) \mid \lambda'(x, y) \geq k\},$$

where

$$\lambda'(x, y) = \left(\frac{p_1}{p_0}\right)^{\sum_1^n b_i} \left(\frac{1 - p_1}{1 - p_0}\right)^{n - \sum_1^n b_i}$$

is the likelihood ratio of the observed b_i's, with

$$p_j = \text{Prob}\{b_i = 1 \mid H_j^*\} = (1 + g\rho_j)^{-1}, \qquad j = 0, 1.$$

The purpose of the present section is to show, in a qualitative and heuristic manner, that under appropriate restrictions

$$\text{Prob}\{W_k \mid H_j^*\} \doteq \text{Prob}\{W_k' \mid H_j^*\} \qquad \textit{for } j = 0, 1$$

and any τ_0, which implies the approximate optimality of a test based on b_i's only, for the given sequential sampling and termination rule. The following discussion could be formulated more quantitatively, but this seems unnecessary since

(a) the tests based on b_i's have simple and known properties,
(b) the operating characteristics of tests based on $\lambda(x, y)$ would apparently be difficult to determine exactly and more difficult to apply than the test on b_i's,
(c) a qualitative indication of the approximate equivalence of the tests serves practically as an additional recommendation for use of the tests based on b_i's.

Now

$$\lambda(x, y) = \frac{(2\pi\rho_1\tau_1)^{-n_x/2}(2\pi\tau_1)^{-n_y/2}\exp\left[-\frac{1}{2\rho_1\tau_1}\sum_1^{n_x} x_i^2 - \frac{1}{2\tau_1}\sum_1^{n_y} y_i^2\right]}{(2\pi\rho_0\tau_0)^{-n_x/2}(2\pi\tau_0)^{-n_y/2}\exp\left[-\frac{1}{2\rho_0\tau_0}\sum_1^{n_x} x_i^2 - \frac{1}{2\gamma_0}\sum_i^{n_y} y_i^2\right]}$$

$$= \left(\frac{\rho_0+1}{\rho_1+1}\right)^{n_x/2}\left(\frac{\rho_1(\rho_0+1)}{\rho_0(\rho_1+1)}\right)^{n_y/2}\exp\left[-\frac{1}{2}\left(\sum_1^{n_x} x_i^2 - \sum_1^{n_y} y_i^2\right)\left(\frac{\rho_0-\rho_1}{\tau_0\rho_0(\rho_1+1)}\right)\right].$$

Now the case $g = 1$ of the basic sampling rule is such that at every stage of sampling the quantity $(\sum x_i^2 - \sum y_i^2)$ will be increased if it is negative and will be decreased if it is positive; thus this quantity will under all hypotheses have a distribution concentrated near the value zero, and the exponential factor of $\lambda(x, y)$ will tend to have a value near unity. For cases $g \neq 1$ of the sampling rule, a change of scale in x_i values, $x_i \to \sqrt{g}x_i$, before computation of $\lambda(x, y)$ gives the same result. We continue the discussion just for the case $g = 1$.

Thus if the rule for termination of sampling is such that $n_x + n_y$ is not small,

$$\text{Prob}\,\{\lambda(x, y) \geqq k \mid H_j^*\} \doteq \text{Prob}\left\{\left(\frac{\rho_0+1}{\rho_1+1}\right)^{n_x/2}\left(\frac{\rho_1(\rho_0+1)}{\rho_0(\rho_1+1)}\right)^{n_y/2} \geqq k \mid H_j^*\right\}$$

and the last quantity is independent of the "nuisance parameter" τ_0, for $j = 0, 1$. From the definition of the b_i's we readily obtain

$$\frac{n_y}{2} - 1 \leqq m - \sum_1^m b_i \leqq \frac{n_y}{2} \quad \text{and} \quad \frac{n_x}{2} - 1 \leqq \sum_1^m b_i \leqq \frac{n_x}{2}.$$

Since

$$p_i = \frac{1}{\rho_i + 1}, \qquad 1 - p_i = \frac{\rho_i}{\rho_i + 1},$$

for $i = 0, 1$, we have

$$\lambda'(x, y) = \left(\frac{p_1}{p_0}\right)^{\Sigma_1^m b_i}\left(\frac{1-p_1}{1-p_0}\right)^{m-\Sigma_1^m b_i}$$

$$\doteq \left(\frac{\rho_0+1}{\rho_1+1}\right)^{n_x/2}\left(\frac{\rho_1(\rho_0+1)}{\rho_0(\rho_1+1)}\right)^{n_y/2}.$$

Hence

$$\text{Prob}\,\{\lambda(x, y) > k \mid H_j^*\} \doteq \text{Prob}\,\{\lambda'(x, y) > k \mid H_j^*\},$$

for $j = 0, 1$, and for each value of the "nuisance parameter" τ_0.

5. Sequential Tests on Components of Variance. The testing problems arising in Model II of the analysis of variance, and their usual non-sequential solution based on F tests, are described, for example, by Mood in [10, Chapter 14). The method of the preceding sections can be adapted to provide sequential tests for such problems, as indicated below. Such sequential tests provide savings like those described above in the required numbers of observations.

Consider the "one-way layout" problem in which

$$y_{ij} = \mu + a_i + e_{ij}, \qquad \text{for } i = 1, 2, \cdots, j = 1, 2, \cdots.$$

Here μ is an unknown constant, a_i and e_{ij} are normally distributed random variables with zero means and unknown respective variances σ_a^2 and σ_e^2, and all a_i's and e_{ij}'s are mutually independent. The statistical problem is to test a hypothesis specifying the value of $\rho = \sigma_a^2 / \sigma_e^2$. For present purposes we consider that a doubly infinite array of the random variables y_{ij} is available, and that we are free to take observations y_{ij} throughout this array in any manner. Let

$$T = (t_{ij}) = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & \cdot \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & -\frac{2}{\sqrt{2\cdot 3}} & 0 & \cdot \\ \vdots & & & & \cdot \\ \cdot & \cdot & \cdot & & \vdots \end{pmatrix}$$

denote the unique "doubly infinite orthogonal matrix" characterized as indicated, i.e., by the requirements that $t_{ij} = 0$ if $j \geq i + 2$, that $t_{i1} > 0$, and that $\sum_{j=1}^{\infty} t_{ij} = 0$, for all i, j. Let the random variables $(r_1, r_2, \cdots)$ be defined by

$$(r_1, r_2, \cdots)' = T(y_{11}, y_{12}, \cdots)':$$

i.e., $r_\alpha = \sum_{j=1}^{\alpha+1} t_{\alpha j} y_{1j}$, for $\alpha = 1, 2, \cdots$. Let the random variables $(s_1, s_2, \cdots)$ be defined by

$$(s_1, s_2, \cdots)' = T(y_{21}, y_{31}, \cdots)';$$

i.e., $s_\alpha = \sum_{j=1}^{\alpha+1} t_{\alpha j} y_{j+1,1}$ for $\alpha = 1, 2, \cdots$. Then all r_α's and s_α's have independent normal distributions with zero means; and $\text{Var}(r_\alpha) = \sigma_r^2 = \sigma_e^2$, $\text{Var}(s_\alpha) = \sigma_s^2 = \sigma_e^2 + \sigma_a^2$.

The sequential test procedures given above can be applied directly to the sequences of r_α's and s_α's to test any hypothesis on $\rho = \sigma_s^2 / \sigma_r^2$, say $H_0: \rho = \rho_0$

against $H_1: \rho = \rho_1 > \rho_0$, at specified size and power. But since

$$\rho = \frac{\sigma_s^2}{\sigma_r^2} = \frac{\sigma_e^2 + \sigma_a^2}{\sigma_e^2} = 1 + \frac{\sigma_a^2}{\sigma_e^2},$$

H_0 is equivalent to $H_0^*: \sigma_a^2/\sigma_e^2 = \rho_0 - 1$, and H_1 is equivalent to $H_1^*: \sigma_a^2/\sigma_e^2 = \rho_1 - 1$. (Here $\rho_0 \geqq 1$ is required if H_0^* is to be meaningful.) Thus the usual hypotheses of interest in terms of variance components, namely those specifying a positive value for $\sigma_a^2/\sigma_e^2 = \rho_0 - 1$, and those specifying $\sigma_a^2 = 0$ (and hence $\rho_0 = 1$), can be tested sequentially with gains in efficiency as described above.

REFERENCES

[1] A. Birnbaum, "Some procedures for comparing Poisson processes or populations", *Biometrika*, Vol. 40 (1953), pp. 447–449.

[2] A. Birnbaum, "Statistical methods for Poisson processes and exponential populations", *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 254–266.

[3] Statistical Research Group, Columbia University, *Selected techniques of statistical analysis*, McGraw-Hill Book Co., Inc., New York, 1947.

[4] National Bureau of Standards, *Tables of the Binomial Probability distribution*, U. S. Government Printing Office, Washington, 1949.

[5] A. Wald, *Sequential Analysis*. John Wiley and Sons, Inc., New York, 1947.

[6] D. Blackwell and M. A. Girshick, *Theory of games and statistical decisions*, John Wiley and Sons, Inc., New York, 1954.

[7] M. A. Girshick, "Contributions to the theory of sequential analysis. I", *Ann. Math. Stat.*, Vol. 17 (1946), pp. 123–143.

[8] S. Rushton, "On sequential tests of the equality of variances of two normal populations with known means," *Sankhyā*, Vol. 12 (1952), pp. 63–78.

[9] N. L. Johnson, "Sequential procedures in certain component of variance problems," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 357–366.

[10] A. Mood, *Intro. to the Theory of Statistics*. McGraw-Hill Book Co., Inc., New York, 1950.

SUMS OF POWERS OF INDEPENDENT RANDOM VARIABLES[1]

By J. M. Shapiro

Ohio State University

1. Summary and introduction. Let (x_{nk}), $k = 1, \cdots, k_n$; $n = 1, 2, \cdots$ be a double sequence of infinitesimal random variables which are rowwise independent (i.e., $\lim_{n\to\infty} \max_{1\leq k\leq k_n} P(|x_{nk}| > \epsilon) = 0$ for every $\epsilon > 0$, and for each n $x_{n1}, \cdots, x_{nk_n}$ are independent). Let $S_n = x_{n1} + \cdots + x_{nk_n} - A_n$ where the A_n are constants and let $F_n(x)$ be the distribution function of S_n. Necessary and sufficient conditions for $F_n(x)$ to converge to a distribution function $F(x)$ are known, and in particular we know that $F(x)$ is infinitely divisible.

In this paper we shall investigate the system of infinitesimal, rowwise independent random variables $(|x_{nk}|^r)$, $r \geq 1$. In particular we shall be interested in large values of r. Specifically, let $S_n^r = |x_{n1}|^r + \cdots + |x_{nk_n}|^r - B_n(r)$, where $B_n(r)$ are suitably chosen constants. Let $F_n^r(x)$ be the distribution function of S_n^r. Necessary and sufficient conditions for $F_n^r(x)$ to converge $(n \to \infty)$ to a distribution function $F^r(x)$ are given, and also necessary and sufficient conditions for $F^r(x)$ to converge $(r \to \infty)$ to a distribution function $H(x)$ are given. The form that $H(x)$ must take is obtained and under rather general conditions it is shown that $H(x)$ is a Poisson distribution. In any case it is shown that $H(x)$ is the sum of two independent random variables, one Gaussian and the other Poisson (including their degenerate cases).

2. Notation. Let $F(x)$ be any infinitely divisible distribution function with characteristic function $\varphi(t)$. According to the formulas of Lévy and Khintchine (cf. [1]) we know that $\varphi(t)$ has the following representation:

$$\varphi(t) = \exp\Big\{i\gamma(\tau)t - \tfrac{1}{2}\sigma^2t^2 + \int_{-\infty}^{-\tau} (e^{itu} - 1)\, dM(u)$$

$$(2.1) \qquad + \int_{\tau}^{\infty} (e^{itu} - 1)\, dN(u) + \int_{-\tau}^{0-} (e^{itu} - 1 - itu)\, dM(u)$$

$$+ \int_{0+}^{\tau} (e^{itu} - 1 - itu)\, dN(u),$$

where $M(u)$ and $N(u)$ are respectively nondecreasing functions in the intervals $(-\infty, 0)$, $(0, +\infty)$ which satisfy $M(-\infty) = N(+\infty) = 0$ and $\int_{-\epsilon}^{0} u^2\, dM(u) + \int_0^{\epsilon} u^2\, dN(u) < \infty$ for every $\epsilon > 0$; σ is a nonnegative constant; τ and $-\tau$ are continuity points of $N(u)$ and $M(u)$; and $\gamma(\tau)$ is a constant depending only on τ.

It is well known that the distribution functions $F^r(x)$ and $H(x)$ referred to in Section 1 are infinitely divisible, and throughout this paper we let $M^r(u)$ and

Received August 28, 1957, revised November 12, 1957.

[1] Presented to the American Mathematical Society August 29, 1957.

$N^r(u)$ be associated with $F^r(x)$ and $M^*(u)$ and $N^*(u)$ be associated with $H(x)$, through the formulas given for their characteristic functions analogous to (2.1).

We let $F_{nk}(x)$ and $F^r_{nk}(x)$ be the distribution functions of x_{nk} and $|x_{nk}|^r$ respectively.

When speaking of a random variable (or its distribution function) being Poisson we shall mean it is either Poisson or its degenerate case (i.e., a random variable taking the value 0 with probability 1). The same applies to a Gaussian random variable) in this case the degenerate case is a random variable taking the value m with probability 1).

If $K(x)$ is a nondecreasing function when we write $\lim_{n\to\infty} K_n(x) = K(x)$ it is understood that this need only hold at continuity points of $K(x)$.

3. General results and proofs.

THEOREM 1. *Let* $\lim_{n\to\infty} F^r_n(x) = F^r(x)$ *for* $r \geqq r_0 \geqq 1$ *and* $\lim_{r\to\infty} F^r(x) = H(x)$, *where* $F^r(x)$ *and* $H(x)$ *are distribution functions. Then* $H(x)$ *is the distribution function of the sum of two independent random variables one of which is Gaussian and the other Poisson.*

We remark that Theorem 1 remains valid if we assume $\lim_{n\to\infty} F^r_n(x) = F^r(x)$ for some sequence of values of r becoming infinite in place of this condition holding for $r \geqq r_0$.

The proof of Theorem 1 requires the following lemma.

LEMMA 1. *If we add to the hypothesis of Theorem* 1 *the condition that* $\lim_{n\to\infty} F_n(x) = F(x)$, *the conclusion of Theorem* 1 *holds.*

Proof. Since $\lim_{n\to\infty} F_n(x) = F(x)$ by Theorem 1 on page 116 of [1], we see

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} F_{nk}(x) = M(x), \qquad x < 0,$$

and

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} (F_{nk}(x) - 1) = N(x), \qquad x > 0,$$

where $M(x)$ and $N(x)$ are given by (2.1). Now for $\alpha \geqq 0$,

$$F^r_{nk}(\alpha) \equiv P(|x_{nk}|^r \leqq \alpha) = F_{nk}(\alpha^{1/r}) - F_{nk}(-\alpha^{1/r}-)$$

and for $\alpha < 0$, $F^r_{nk}(\alpha) = 0$. Thus for $x < 0$, $\lim_{n\to\infty} \sum_{k=1}^{k_n} F^r_{nk}(x) = 0$, and for $x > 0$,

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} [F^r_{nk}(x) - 1] = \lim_{n\to\infty} \sum_{k=1}^{k_n} [F_{nk}(x^{1/r}) - 1] + \lim_{n\to\infty} \sum_{k=1}^{k_n} [-F_{nk}(-x^{1/r}-)].$$

Now assume that $x^{1/r}$ and $-x^{1/r}$ are continuity points of $N(x)$ and $M(x)$ respectively. Note that the set of points $x > 0$ for which this is true is dense on the positive axis. For such x we have $\lim_{n\to\infty} \sum_{k=1}^{k_n} [F^r_{nk}(x) - 1] = N(x^{1/r}) -$

$M(-x^{1/r})$. We note that the function $N(x^{1/r}) - M(-x^{1/r})$ and the function

$$\sum_{k=1}^{k_n} [F^r_{nk}(x) - 1]$$

are both nondecreasing for $x > 0$ and hence $\lim_{n\to\infty} \sum_{k=1}^{k_n} [F^r_{nk}(x) - 1] = N(x^{1/r}) - M(-x^{1/r})$ at all continuity points, $x > 0$, of $N(x^{1/r}) - M(-x^{1/r})$. Now since $\lim_{n\to\infty} F^r_n(x) = F^r(x)$ we see by Theorem 1 on page 116 of [1] that $M^r(x) \equiv 0$ and $N^r(x) = N(x^{1/r}) - M(-x^{1/r})$ (Note that since $\int_{-\epsilon}^{0} x^2\, dM(x) + \int_0^{\epsilon} x^2\, dN(x) < \infty$ it follows that for $r \geqq 1$, $\int_{-\epsilon}^{0} x^2\, dM^r(x) + \int_0^{\epsilon} x^2\, dN^r(x) < \infty$.) Now since $\lim_{r\to\infty} F^r(x) = H(x)$, it follows by Theorem 2 on page 88 of [1] that $\lim_{r\to\infty} M^r(x) = M^*(x)$ and $\lim_{r\to\infty} N^r(x) = N^*(x)$ at continuity points of $M^*(x)$ and $N^*(x)$. This shows that $M^*(x) \equiv 0$ and

$$N^*(x) = \lim_{r\to\infty} [N(x^{1/r}) - M(-x^{1/r})] = \begin{cases} N(1+) - M(-1-), & x > 1, \\ N(1-) - M(-1+), & 0 < x < 1. \end{cases}$$

This shows that $N^*(x)$ is constant for $0 < x < 1$ and for $x > 1$ and hence (since $M^*(-\infty) = N^*(+\infty) = 0$) we see that $N(1+) = 0$ and $M(-1-) = 0$ Thus we see $N^*(x)$ is either identically 0 or takes one jump at $x = 1$. (In fact if both $M(x)$ is continuous at -1, $N(x)$ is continuous at $+1$ then $N^*(x) = 0$; otherwise $N^*(x)$ takes one jump). Now let σ^* be the nonnegative constant associated with $H(x)$ by the formula (2.1). Then if $\sigma^* = 0$ and $N^*(x)$ takes one jump it is clear that $H(x)$ is Poisson or $H(x - m)$ is Poisson (m a constant). If $\sigma^* = 0$ and $N^*(x) \equiv 0$, $H(x)$ is a unitary distribution. If $\sigma^* = 0$ and $N^*(x) \equiv 0$, $H(x)$ is Gaussian; and if $\sigma^* = 0$ and $N^*(x)$ takes one jump, then (cf. [1]) it follows that $H(x)$ is the sum of two independent random variables one Gaussian and the other Poisson. This proves the lemma

Proof of Theorem 1. Let $s \geqq r_0$ and let $y_{nk} = |x_{nk}|^s$. Then $|x_{nk}|^r = |y_{nk}|^{r/s}$. Then for $r/s \geqq 1$, under the conditions of Theorem 1 the conditions of Lemma 1 are satisfied with the system (x_{nk}) replaced by (y_{nk}). This proves Theorem 1.

LEMMA 2. *If* $\lim_{n\to\infty} F_n(x) = F(x)$, *then for suitably chosen constants* $B_n(r)$, $F^r_n(x)$ *converges to a distribution function* $F^r(x)$ *if and only if*[3]

$$\lim_{\epsilon\to 0} \lim_{n\to\infty}{}^{\sup}_{\inf} \sum_{k=1}^{k_n} \left\{ \int_0^{\epsilon} x^{2r}\, d[F_{nk}(x) - F_{nk}(-x-)] - \left(\int_0^{\epsilon} x^r\, d[F_{nk}(x) - F_{nk}(-x-)] \right)^2 \right\} = \sigma_r^2 < \infty. \tag{3.1}$$

Proof. Suppose $\lim_{n\to\infty} F_n(x) = F(x)$ and that (3.1) holds. Then as in the proof Lemma 1, $\lim_{n\to\infty} \sum_{k=1}^{k_n} F^r_{nk}(x) = 0 = M^r(x)$ for $x < 0$, and

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} (F^r_{nk}(x) - 1) = N(x^{1/r}) - M(-x^{1/r}) = N^r(x) \qquad \text{for } x > 0$$

[3] We use the notation $\lim{}^{\sup}_{\inf}$ to mean that the indicated condition is to hold for both lim inf and lim sup.

(Here we consider $N^r(x)$ and $M^r(x)$ only as functions just defined and not, at this point, as being associated with any distribution function.) We see that $M^r(-\infty) = N^r(+\infty) = 0$ and that $\int_{-\epsilon}^{0} x^2\, dM^r(x) + \int_0^{\epsilon} x^2\, dN^r(x) < \infty$. Consider

$$\lim_{\epsilon\to 0} \lim_{n\to\infty}{}^{\sup}_{\inf} \sum_{k=1}^{k_n} \left\{ \int_{|x|<\epsilon} x^2\, dF_{nk}^r(x) - \left(\int_{|x|<\epsilon} x\, dF_{nk}^r(x) \right)^2 \right\}$$

$$= \lim_{\epsilon\to 0} \lim_{n\to\infty}{}^{\sup}_{\inf} \sum_{k=1}^{k_n} \left\{ \int_0^{\epsilon} x^2\, d[F_{nk}(x^{1/r}) - F_{nk}(-x^{1/r}-)] \right.$$

$$\left. - \left(\int_0^{\epsilon} x\, d[F_{nk}(x^{1/r}) - F_{nk}(-x^{1/r}-)] \right)^2 \right\} \tag{3.2}$$

$$= \lim_{\epsilon\to 0} \lim_{n\to\infty}{}^{\sup}_{\inf} \sum_{k=1}^{k_n} \left\{ \int_0^{\epsilon^{1/r}} x^{2r}\, d[F_{nk}(x) - F_{nk}(-x-)] \right.$$

$$\left. - \left(\int_0^{\epsilon^{1/r}} x^r\, d[F_{nk}(x) - F_{nk}(-x-)] \right)^2 \right\} = \sigma_r^2,$$

using condition (3.1). (Note r is fixed here.) Now by choosing

$$B_n(r) = \sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF_{nk}^r(x) - C_r + o\,(1),$$

where C_r is a constant and $o(1) \to 0$ as $n \to \infty$, we see by Theorem 1 on page 116 of [1] that $F_n^r(x)$ converges to a distribution $F^r(x)$. (We note that $M^r(x)$, $N^r(x)$ and σ_r^2 are associated with $F^r(x)$ through the formulas (2.1).)

Now suppose that $F_n^r(x) \to F^r(x)$. Then again using the theorem of [1] referred to above we see that (3.2) holds and hence that (3.1) holds.

THEOREM 2. *Under the conditions of Lemma 2 a necessary and sufficient condition for the distribution functions $F^r(x)$ to converge $(r \to \infty)$ to a distribution function $H(x)$ for suitably chosen constants $B_n(r)$, is that*[3]

$$M(x) = 0 \quad \text{for} \quad x < -1, \qquad N(x) = 0 \quad \text{for} \quad x > 1,$$
$$\lim_{r\to\infty} \sigma_r^2 = (\sigma^*)^2. \tag{3.3}$$

Furthermore $H(x)$ is Gaussian if $M(x)$ is continuous at -1 and $N(x)$ is continuous at $+1$, $H(x - m)$ is Poisson if $\sigma^* = 0$ and either $M(x)$ is discontinuous at -1 or $N(x)$ is discontinuous at $+1$ where m is a constant, and $H(x)$ is the sum of two independent random variables, one Gaussian and the other Poisson otherwise.

Proof. Suppose $\lim_{r\to\infty} F^r(x) = H(x)$. Then as in the proof of Lemma 1 we see that $M(-1-) = 0$ and $N(1+) = 0$ and hence $M(x) = 0$ for $x < -1$ and $N(x) = 0$ for $x > 1$. Now by Theorem 2 on page 88 of [1] we have

$$\lim_{\epsilon\to 0} \lim_{r\to\infty}{}^{\sup}_{\inf} \left\{ \int_{-\epsilon}^{0} u^2\, dM^r(u) + \sigma_r^2 + \int_0^{\epsilon} u^2\, dN^r(u) \right\} = (\sigma^*)^2.$$

---

[3] Same notation as in the proofs of Lemmas 1 and 2.

Now

$$\left\{\int_{-\epsilon}^{0} u^2\, dM^r(u) + \int_0^{\epsilon} u^2\, dN^r(u)\right\} = \int_0^{\epsilon} u^2\, d[N(u^{1/r}) - M(-u^{1/r})]$$

$$= \int_0^{\epsilon^{1/r}} y^{2r}\, d[N(y) - M(-y)]$$

$$\leqq \epsilon\left\{\int_0^1 y^2\, dN(y) + \int_{-1}^{0} y^2\, dM(y)\right\} \qquad \text{for } r > 1 \text{ and } 0 < \epsilon < 1.$$

Thus we see that $\liminf_{r\to\infty} \sigma_r^2 = (\sigma^*)^2 = \limsup_{r\to\infty} \sigma_r^2$.

Now suppose (3.3) holds Then as in the proof of Lemma 1 we see

$$\lim_{r\to\infty} N^r(x) = N^*(x) = \begin{cases} N(1+) - M(-1-) & \text{for } x > 1 \\ N(1-) - M(-1+) & \text{for } 0 < x < 1 \end{cases}$$

and $\lim_{r\to\infty} M^r(x) = 0 \equiv M^*(x)$. (Here we consider M^* and N^* as functions just defined and not at this point as being associated with $H(x)$.) Now from (3.3) it follows that $N^*(+\infty) = M^*(-\infty) = 0$ and $\int_{-\epsilon}^{0} x^2\, dM^*(x) + \int_0^{\epsilon} x^2\, dN^*(x) < \infty$. Also since

$$\lim_{\epsilon\to 0} \lim_{r\to\infty}{}^{\sup}_{\inf} \left\{\int_{-\epsilon}^{0} u^2\, dM^r(u) + \int_0^{\epsilon} u^2\, dN^r(u)\right\} = 0$$

(from the first part of this proof), it follows from (3.3) that

$$\lim_{\epsilon\to 0} \lim_{r\to\infty}{}^{\sup}_{\inf} \left\{\int_{-\epsilon}^{0} u^2\, dM^r(u) + \sigma_r^2 + \int_0^{\epsilon} u^2\, dN^r(u)\right\} = (\sigma^*)^2.$$

Now by Theorem 1 on page 116 of [1] we see that $\gamma_r(\tau) = \sum_{k=1}^{k_n} \int_{|x|<\tau} x\, dF^r_{nk}(x) - B_n(r) + o(1)$, where $\gamma_r(\tau)$ is associated with $F^r(x)$ through the formulas (2.1). Thus by the proper choice of $B_n(r)$, $\gamma_r(\tau)$ converges $(r \to \infty)$ to some constant $\gamma_*(\tau)$, (τ fixed). But using Theorem 2 on page 88 of [1], we see that $\lim_{r\to\infty} F^r(x) = H(x)$, where $H(x)$ is the infinitely divisible distribution determined by M^*, N^*, $\gamma_*(\tau)$ and $(\sigma^*)^2$ given above. *It remains to show the form for $H(x)$, but this* follows as in the proof of Lemma 1.

4. Characterization of the Poisson distribution. In this section we give conditions which will insure that the distribution functions $F^r(x)$ will converge to the Poisson distribution. We use the same notation as in the previous sections. In particular $M(x)$ and $N(x)$ are associated with the distribution function $F(x)$, the limiting distribution of $F_n(x)$.

THEOREM 3. *If $F_n(x)$ converges to $F(x)$, $M(x) = 0$ for $x < -1$, $N(x) = 0$ for $x > 1$, and*

$$(4.1) \qquad \sum_{k=1}^{k_n} \int_{|x|<\epsilon} |x|^s\, dF_{nk}(x) \quad \textit{is bounded in } n \textit{ for some } s < 2r,$$

then for suitably chosen constants $B_n(r)$, $F_n^r(x)$ *converges* ($n \to \infty$) *to a distribution function* $F^r(x)$ *and* $F^r(x)$ *converges* ($r \to \infty$) *to the Poisson distribution. (No matter what the choice of* $B_n(r)$, *if* $F_n^r(x) \to F^r(x)$ *and* $F^r(x) \to H(x)$, *then there exists a constant* m *such that* $H(x - m)$ *is Poisson.)*

We postpone the proof of Theorem 3 as well as that of the next three theorems. In the rest of the paper it will be convenient to assume $r > 1$.

THEOREM 4. *Condition* (4.1) *of Theorem* 3 *may be replaced by*

(4.2) *The random variables* (x_{nk}) *are symmetric about the origin.*

THEOREM 5. *Let* $S_n = x_{n1} + \cdots + x_{nk_n}$ (i.e., *let* $A_n = 0$) *and suppose* $F_n(x) \to F(x)$. *Let* $N(x) = 0$ *for* $x > 1$. *Then if the* (x_{nk}) *are positive random variables the conclusion of Theorem* 3 *holds.*

THEOREM 6. *Let* $A_n = 0$, $F_n(x) \to F(x)$, $M(x) = 0$ *for* a < -1, *and* $N(x) = 0$ *for* $x > 1$. *Then if the* (x_{nk}) *are identically distributed within each row the conclusion of Theorem* 3 *holds.*

Proof of Theorem 3. We first show that condition (4.1) implies condition (3.1) of Lemma 2 with $\sigma_r^2 = 0$. We have

$$\sum_{k=1}^{k_n} \left\{ \int_0^\epsilon x^{2r}\, d[F_{nk}(x) - F_{nk}(-x-)] - \left(\int_0^\epsilon x^r\, d[F_{nk}(x) - F_{nk}(-x-)] \right)^2 \right\}$$

$$\leqq \sum_{k=1}^{k_n} \left\{ \int_0^\epsilon x^{2r}\, d[F_{nk}(x) - F_{nk}(-x-)] \right\}$$

$$\leqq \epsilon^{2r-s} \sum_{k=1}^{k_n} \left\{ \int_0^\epsilon |x|^s\, d[F_{nk}(x) - F_{nk}(-x-)] \right.$$

$$= \epsilon^{2r-s} \sum_{k=1}^{k_n} \left\{ \int_0^\epsilon |x|^s\, dF_{nk}(x) + \int_{-\epsilon}^0 |x|^s\, dF_{nk}(x-) \right\},$$

and since $2r - s > 0$ we see by (4.1) that (3.1) holds with $\sigma_r^2 = 0$. Thus from Lemma 2, $F_n^r(x) \to F^r(x)$. Also since $\lim_{r\to\infty} \sigma_r^2 = 0 = (\sigma^*)^2$, it follows from Theorem 2 that $F^r(x) \to H(x)$ and that $H(x - m)$ is a Poisson distribution. (This includes the possibility that $H(x)$ may be a degenerate Gaussian distribution.) We note that $B_n(r)$ could be chosen so as to make $m = 0$. This proves the theorem.

Proof of Theorem 4. We only need to show that (4.2) implies (4.1). Let $\alpha_{nk} = \int_{|x|<\tau} x\, dF_{nk}(x)$ for some $\tau > 0$. By Theorem 2 on page 111 of [1] we have

$$\sum_{k=1}^{k_n} \int_{-\infty}^{\infty} \frac{x^2}{1 + x^2}\, dF_{nk}(x + \alpha_{nk})$$

is bounded. But since the random variables are symmetric it follows that $\alpha_{nk} = 0$ and hence

$$\sum_{k=1}^{k_n} \int_{|x|<\epsilon} x^2 \, dF_{nk}(x) \leq (1 + \epsilon^2) \sum_{k=1}^{k_n} \int_{|x|<\epsilon} \frac{x^2}{1+x^2} \, dF_{nk}(x)$$

$$\leq (1 + \epsilon^2) \sum_{k=1}^{k_n} \int_{-\infty}^{\infty} \frac{x^2}{1+x^2} \, dF_{nk}(x)$$

is bounded. Thus (4.1) holds with $s = 2$, i.e., for $r > 1$.

Proof of Theorem 5. Since the x_{nk} are positive it follows from Theorem 1 on page 116 of [1] that $M(x) \equiv 0$ for $x < 0$, and that $\sum_{k=1}^{k_n} \int_{|x|<\tau} x \, dF_{nk}(x) = \sum_{k=1}^{k_n} \int_0^{\tau} x \, dF_{nk}(x)$ converges to a constant $\gamma(\tau)$ (note $A_n = 0$). Thus

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \left[\int_{|x|<\epsilon} x \, dF_{nk}(x) \right]^2 = \lim_{n\to\infty} \sum_{k=1}^{k_n} \left[\int_0^{\epsilon} x \, dF_{nk}(x) \right]^2$$

$$\leq \lim_{n\to\infty} \left[\max_{1 \leq k \leq k_n} \left(\int_0^{\epsilon} x \, dF_{nk}(x) \right) \right] \left[\lim_{n\to\infty} \sum_{k=1}^{k_n} \int_0^{\epsilon} x \, dF_{nk}(x) \right] = 0,$$

since $\lim_{n\to\infty} \max_{1\leq k \leq k_n} \int_0^{\epsilon} x \, dF_{nk}(x) = 0$ (infinitesimalness). Now again from Theorem 1 on page 116 of [1] we have

$$\lim_{\epsilon\to 0} \lim{}^{\sup}_{\inf}{}_{n\to\infty} \sum_{k=1}^{k_n} \left\{ \int_{|x|<\epsilon} x^2 \, dF_{nk}(x) - \left(\int_{|x|<\epsilon} x \, dF_{nk}(x) \right)^2 \right\} = \sigma^2$$

so that

$$\lim_{\epsilon\to 0} \lim{}^{\sup}_{\inf}{}_{n\to\infty} \sum_{k=1}^{k_n} \int_{|x|<\epsilon} x^2 \, dF_{nk}(x) = \sigma^2 < \infty.$$

Thus $\sum_{k=1}^{k_n} \int_{|x|<\epsilon} x^2 \, dF_{nk}(x)$ is bounded in n so that (4.1) holds with $s = 2$. This proves Theorem 5.

Proof of Theorem 6. Since $A_n = 0$ we again have

$$\sum_{k=1}^{k_n} \int_{|x|<\tau} x \, dF_{nk}(x) = k_n \int_{|x|<\tau} x \, dF_{n1}(x) \to \gamma(\tau).$$

Also

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \left(\int_{|x|<\epsilon} x \, dF_{nk}(x) \right)^2 = \lim_{n\to\infty} k_n \left(\int_{|x|<\epsilon} x \, dF_{n1}(x) \right)^2$$

$$= \lim_{n\to\infty} k_n \int_{|x|<\epsilon} x \, dF_{n1}(x) \cdot \lim_{n\to\infty} \int_{|x|<\epsilon} x \, dF_{n1}(x)$$

$$= \gamma(\tau) \cdot 0 = 0,$$

since the random variables (x_{nk}) are infinitesimal. From this point the proof is identical to that of Theorem 5.

The next theorem shows the existence of a double sequence of random variables $(|x_{nk}|^r)$ such that the distribution functions of the row sums (minus a constant) converge to the Poisson distribution.

THEOREM 7. *Under the conditions of any one of the Theorems 3 through 6 there exists a sequence of numbers $r_n \to \infty$ such that the distribution functions of the sums $|x_{n1}|^{r_n} + \cdots + |x_{nk_n}|^{r_n} - B_n(r_n)$, ($B_n(r_n)$ suitably chosen constants) converge to the Poisson distribution.*[4]

Proof. We have $\lim_{n\to\infty} F_n^r(x) = F^r(x)$ and $\lim_{r\to\infty} F^r(x) = H(x)$, where $H(x)$ is a Poisson distribution. (In particular the first limit relation holds for $r = 2, 3, \cdots$.) Let $\{\xi_k\}$, $k = 1, 2, \cdots$, be a countable dense set on the real line such that $F_n^r(\xi_k) \to F^r(\xi_k)$ for $r = 2, 3, \cdots$ and $F^r(\xi_k) \to H(\xi_k)$ for all k. Let $\{\epsilon_n\}$ be a positive decreasing sequence of real numbers such that $\epsilon_n \to 0$ as $n \to \infty$. Let $\{n_r\}$ be an increasing subsequence of the positive integers such that $n \geqq n_r$ implies that $|F_n^r(\xi_k) - F^r(\xi_k)| < \epsilon_r$ for $k = 1, 2, \cdots, r$ (r fixed). Consider the sequence of distribution functions S: $F_1^2(x)$, $F_2^2(x)$, $\cdots$, $F_{n_3-1}^2(x)$, $F_{n_3}^3(x)$, $\cdots$, $F_{n_4-1}^3(x)$, $F_{n_4}^4(x)$, $\cdots F_{n_5-1}^4(x)$, $\cdots$. We claim this sequence converges to $H(x)$ for $x = \xi_k$ for $k = 1, 2, \cdots$. Consider ξ_k. Let $\epsilon > 0$ be given. Let r_0, ($r_0 > k$) be such that $r \geqq r_0$ implies $\epsilon_r < \epsilon/2$. Let $r_1 \geqq r_0$ be such that $r \geqq r_1$ implies $|F^r(\xi_k) - H(\xi_k)| < \epsilon/2$. Then for $n > N(\xi_k) = n_{r_1}$ consider

$$|F_u^r(\xi_k) - H(\xi_k)| \leqq |F_n^r(\xi_k) - F^r(\xi_k)| + |F^r(\xi_k) - H(\xi_k)|.$$

Since we are considering only elements of the sequence S we have $n > n_{r_1}$ implies $r \geqq r_1 \geqq r_0 > k$. Therefore $|F_n^r(\xi_k) - F^r(\xi_k)| < \epsilon_r < \epsilon/2$ and $|F^r(\xi_k) - H(\xi_k)| < \epsilon/2$. Thus $|F_n^r(\xi_k) - H(\xi_k)| < \epsilon$ and we see that the sequence S converges to $H(x)$ for $x = \xi_k$, $k = 1, 2, \cdots$. But since $\{\xi_k\}$ is dense, the sequence S converges to $H(x)$ at every continuity point of $H(x)$. Now if we let $r_n = 2$ for $n = 1, \cdots, n_3 - 1$ and $r_n = m$ for $n = n_m, \cdots, n_{m+1} - 1$ ($m > 2$), we see that the distribution function of $|x_{n1}|^{r_n} + \cdots + |x_{nk_n}|^{r_n} - B_n(r_n)$ is $F_n^{r_n}(x)$, which is the nth element of the sequence S. This proves the theorem.

REFERENCE

[1] B. V. GNEDENKO AND A. N. KOLMOGOROV, *Limit Distributions for Sums of Independent Random Variables*, translation by K. L. Chung Addison-Wesley, 1954.

[4] An analogous theorem holds for the conditions of theorem 2.

THE STAIRCASE DESIGN: THEORY

By Franklin A. Graybill and William E. Pruitt

Oklahoma State University

0. Summary and Introduction. One of the most popular designs in experimental work is the randomized block. These designs can be put into three broad classes viz. complete block design, balanced incomplete block design, and the partially balanced incomplete block design. These designs are all special cases of the general *two way classification with unequal numbers in the* subclasses, but since the analysis of this general classification is quite complex, these special cases have evolved which are adequate to fit most needs and the analysis of these special designs is relatively easy. [1], [2], [6], [8].

However, most of the block designs considered to date have one feature in common—they require each block to contain an equal number of experimental units. The exceptions are given in [9], [10], where designs are considered in which the number of experimental units in blocks differ by one. The purpose of this paper is to extend the randomized block design to include the case where all blocks do not contain the same number of experimental units. We have called this the *staircase design*

Suppose an experimenter, wishing to run an experiment using N treatments, decides to use a randomized block design, but after arranging his material into homogeneous groups he finds that he has blocks available which have varying number of experimental units. The experimenter has various courses open to him: (1) If enough blocks are available with N or more experimental units he can discard the extra units in these blocks, discard all the blocks which have less than N units, and use a randomized complete block design; (2) He can discard units in the blocks until he has enough units and blocks *for a balanced incomplete block or a partially balanced incomplete block design*; (3) He can use all the experimental units and use the staircase design proposed in this paper.

For example, if an experimenter has N treatments with which he wishes to experiment using a randomized block design, and if he has blocks of unequal size, then he must rank his N treatments in the order of their importance, i.e., $T_1, T_2, \cdots, T_N$, where he considers T_1 the most important and T_N the least important. Now suppose he has at his disposal b_1 blocks which each contain N experimental units. Then all N treatments are randomized in each of the b_1 blocks. Suppose further that he has b_2 blocks which each contain N_1 experimental units ($N_1 < N$). Then the first N_1 treatments are arranged at random in each of the b_2 blocks. This process is continued until all the blocks are used.

A particular example where this would be useful is an experiment involving animals as experimental units where a block consists of litter mates. Let us

Received July 16, 1956; revised December 26, 1957.

suppose that we have two litters of size seven, three of size five, and one of size four. Using the staircase design we can include seven treatments and still have the four we are most interested in replicated six times.

1. Notation. Consider a two-way classification model

$$(1.1) \qquad Y_{ij} = \mu + \beta_i + \alpha_j + e_{ij}; \qquad i = 1, 2, \cdots, c_j; j = 1, 2, \cdots, N.$$

where μ, β_i, α_j are constants and the e_{ij} are normal independent variables with means zero and variances σ^2. Also the j's will be ordered in such a way that $c_j \geqq c_{j'}$ for $j < j'$. The purpose of this paper is,

1. to derive the least squares method for testing the hypothesis $\alpha_1 = \alpha_2 = \cdots = \alpha_N$ under the model given above and to give the power function of this test.
2. to derive the best, linear, unbiased estimates for $\alpha_j - \alpha_{j'}$, and the variances of these estimates.

First we will separate the j's into subsets such that j and j' will be in the same subset if and only if $c_j = c_{j'}$. Each of these subsets will be called a *step*. We will designate the number of steps as k.

Let

$$c_j = M^1 \quad \text{for} \quad j = 1, 2, \cdots, N_1$$
$$c_j = M^2 \quad \text{for} \quad j = N_1 + 1, N_1 + 2, \cdots, N_1 + N_2,$$
$$\vdots \qquad \vdots$$
$$c_j = M^k \quad \text{for} \quad j = N_1 + N_2 + \cdots + N_{k-1} + 1, N_1 + N_2 + \cdots + N_{k-1} + 2, \cdots, N_1 + N_2 + \cdots + N_k,$$

where

$$\sum_{t=1}^{k} N_t = N; \qquad \sum_{t=1}^{k} M^t N_t = N^*$$

Now, let

$$N^p = \sum_{t=1}^{p} N_t, \qquad N^0 = 0, \qquad M^{k+1} = 0,$$
$$Y_{ij} = Y_{ij}^s \quad \text{for} \quad i = 1, 2, \cdots, M^s, j = N^{s-1} + 1, N^{s-1} + 2, \cdots, N^s,$$
$$Y_{ij} = Y_{ij}'^s \quad \text{for} \quad i = 1, 2, \cdots, M^{s+1}, j = 1, 2, \cdots, N^s$$
$$\alpha_j = \alpha_j^s \quad \text{for} \quad j = N^{s-1} + 1, N^{s-1} + 2, \cdots N^s,$$
$$\alpha_j = \alpha_j'^s \quad \text{for} \quad j = 1, 2, \cdots, N^s.$$

Fig. 1 will serve to illustrate some of the notation. It will be noticed that c_j is the number of blocks in which treatment j appears; M^t is the number of blocks in the tth step; N_t is the number of treatments in the tth step. Also Y_{ij}^s is the observation of the jth treatment which appears in the ith block of the sth step. It may be helpful to note further that $Y_{ij}'^1$ is a subset of Y_{ij}^1; $Y_{ij}'^2$ is a

subset of the union of Y_{ij}^1 and Y_{ij}^2, $Y_{ij}'^3$ is a subset of the union of Y_{ij}^1 and Y_{ij}^2 and Y_{ij}^3, etc.

A subscript replaced by a dot indicates the mean of the elements when summed over the range of the replaced subscript, eg.

$$Y_{..}'^2 = \frac{\sum_{i=1}^{M^2} \sum_{j=1}^{N^2} Y_{ij}'^2}{M^2 N^2}.$$

Since superscripts are being used in abundance, a Y, M, N, or α that is raised to a power will always be enclosed in the appropriate brackets.

If, in a summation, the lower limit of summation should exceed the upper limit of summation, the sum will be zero.

The notation used in Section 3 is that used by Kempthorne [4], pages 79–82, with the following exceptions. To be consistent with the notation given above, the normal equations are divided by a constant to give them in terms of means instead of totals. Q_j^s will refer to only the Q_j where $j = N^{s-1} + 1, N^{s-1} + 2, \cdots, N^s$.

2. The test function and its distributional properties. The purpose of this section is to give a test of the hypothesis $\alpha_1 = \alpha_2 = \cdots = \alpha_V$ and to derive the

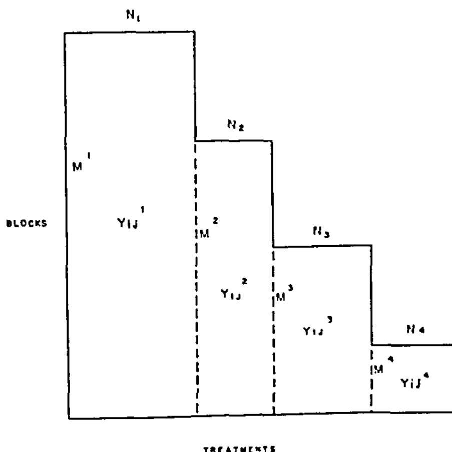


FIG 1

distributional properties of the test function. The proof that this test is the same as that given by the method of least squares will be given in the next section.

Consider the following quadratic forms:

$$q_t^1 = \sum_{i=1}^{M^t} \sum_{j=N^{t-1}+1}^{N^t} (Y_{ij}^t - Y_{i.}^t - Y_{.j}^t + Y_{..}^t)^2, \qquad t = 1, 2, \cdots, k.$$

$$q_t^2 = \frac{N^t N_{t+1}}{N^{t+1}} \sum_{i=1}^{M^{t+1}} (Y_{i.}'^t - Y_{i.}^{t+1} - Y_{..}'^t + Y_{..}^{t+1})^2, \qquad t = 1, 2, \cdots, k-1,$$

$$q_t^3 = M^t \sum_{j=N^{t-1}+1}^{N^t} (Y_{.j}^t - Y_{..}^t)^2, \qquad t = 1, 2, \cdots, k.$$

$$q_t^4 = \frac{M^{t+1} N^t N_{t+1}}{N^{t+1}} (Y_{..}'^t - Y_{..}^{t+1})^2, \qquad t = 1, 2, \cdots, k-1.$$

$$q_t^5 = \frac{1}{N^t} \sum_{i=M^{t+1}+1}^{M^t} (N^{t-1} Y_{i.}'^{t-1} + N_t Y_{i.}^t)^2, \qquad t = 1, 2, \cdots, k.$$

$$q_t^6 = \sum_{i=1}^{M^t} \sum_{j=N^{t-1}+1}^{N^t} (Y_{ij}^t)^2, \qquad t = 1, 2, \cdots, k$$

We will prove the following:

THEOREM I. *If*

$$v = \frac{\sum_{t=1}^k q_t^3 + \sum_{t=1}^{k-1} q_t^4}{\sum_{t=1}^k q_t^1 + \sum_{t=1}^{k-1} q_t^2} \cdot \frac{(M^1 - 1)(N - 1) - \sum_{t=2}^k (M^1 - M^t)(N_t)}{N - 1} \tag{2.1}$$

then v is distributed as $F'_{p,q,\lambda}$, where $F'_{p,q,\lambda}$ represents the non-central F with degrees of freedom p and q and non-centrality λ [7], *also*

$$\begin{aligned} p &= N - 1, \; q = (M^1 - 1)(N - 1) - \sum_{t=2}^k (M^1 - M^t) N_t, \\ \lambda &= \sum_{t=1}^k \left[\frac{M^t}{2\sigma^2} \sum_{j=N^{t-1}+1}^{N^t} (\alpha_j^t - \alpha_.^t)^2\right] + \sum_{t=1}^{k-1} \left[\frac{M^{t+1} N^t N_{t+1}}{2\sigma^2 N^{t+1}} (\alpha_.'^t - \alpha_.^{t+1})^2\right] \end{aligned} \tag{2.2}$$

and $\lambda = 0$ if and only if $\alpha_1 = \alpha_2 = \cdots = \alpha_N$.

PROOF. It is clear that

$$\sum_{t=1}^k q_t^1 + \sum_{t=1}^{k-1} q_t^2 + \sum_{t=1}^k q_t^3 + \sum_{t=1}^{k-1} q_t^4 + \sum_{t=1}^k q_t^5 = \sum_{t=1}^k q_t^6. \tag{2.3}$$

Now it is easily shown that the rank of q_t^1 is $(M^t - 1)(N_t - 1)$, the rank of q_t^2 is $(M^{t+1} - 1)$, the rank of q_t^3 is $(N_t - 1)$, the rank of q_t^4 is 1, and the rank q_t^5 is $(M^t - M^{t+1})$. Adding we see that

$$\sum_{t=1}^k (M^t - 1)(N_t - 1) + \sum_{t=1}^{k-1} (M^{t+1} - 1) + \sum_{t=1}^k (N_t - 1) + (k-1) + \sum_{t=1}^{k-1} (M^t - M^{t+1}) + (M^k - M^{k+1}) = \sum_{t=1}^k M^t N_t$$

Thus we have the fact that the sum of the ranks of the quadratic forms on the left of (2.3) above is equal to the number of squared observations on the right. We may now invoke a theorem proved by Madow [5] showing the quadratic forms to be independent, and verifying the following distributions. (E will be used to denote mathematical expectation, and $\chi'^2_{p,\lambda}$ will represent a non-central chi square distribution with degrees of freedom p and non-centrality λ).

1. q_i^1/σ^2 is distributed as $\chi'^2_{p,\lambda}$, where $p = (M^i - 1)(N_i - 1)$, $\lambda = 0$, since

$$E(Y^i_{ij} - Y^i_{i\cdot} - Y^i_{\cdot j} + Y^i_{\cdot\cdot}) = 0.$$

2. q_i^2/σ^2 is distributed as $\chi'^2_{p,\lambda}$, where $p = (M^{i+1} - 1)$, $\lambda = 0$, since

$$E(Y'^i_{i\cdot} - Y^{i+1}_{i\cdot} - Y'^i_{\cdot\cdot} + Y^{i+1}_{\cdot\cdot}) = 0.$$

3. q_i^3/σ^2 is distributed as $\chi'^2_{p,\lambda}$, where $p = (N_i - 1)$,

$$\lambda = \frac{M^i}{2\sigma^2} \sum_{j=N^{i-1}+1}^{N^i} (\alpha^i_j - \alpha^i_\cdot)^2,$$

since

$$E(Y^i_{\cdot j} - Y^i_{\cdot\cdot}) = \alpha^i_j - \alpha^i_\cdot .$$

4. q_i^4/σ^2 is distributed as $\chi'^2_{p,\lambda}$, where $p = 1$,

$$\lambda = \frac{M^{i+1}N^iN_{i+1}}{2\sigma^2 N^{i+1}} (\alpha'^i_\cdot - \alpha^{i+1}_\cdot)^2,$$

since

$$E(Y'^i_{\cdot\cdot} - Y^{i+1}_{\cdot\cdot}) = \alpha'^i_\cdot - \alpha^{i+1}_\cdot.$$

Therefore it follows that

$$\frac{1}{\sigma^2}\left[\sum_{i=1}^{k} q_i^1 + \sum_{i=1}^{k-1} q_i^2\right]$$

is distributed as $\chi'^2_{p,\lambda}$, where

$$p = (M^1 - 1)(N - 1) - \sum_{i=2}^{k} (M^1 - M^i)(N_i),$$

and $\lambda = 0$. Also we have

$$\frac{1}{\sigma^2}\left[\sum_{i=1}^{k} q_i^3 + \sum_{i=1}^{k-1} q_i^4\right]$$

is distributed as $\chi'^2_{p,\lambda}$, where

$$p = \sum_{i=1}^{k} (N_i - 1) + (k - 1) = N - 1,$$

$$\lambda = \sum_{i=1}^{k}\left[\frac{M^i}{2\sigma^2} \sum_{j=N^{i-1}+1}^{N^i} (\alpha^i_j - \alpha^i_\cdot)^2\right] + \sum_{i=1}^{k-1} \frac{M^{i+1}N^iN_{i+1}}{2\sigma^2 N^{i+1}} (\alpha'^i_\cdot - \alpha^{i+1}_\cdot)^2$$

TABLE 3.1

Due to	df	Sum of Squares
Blocks ignoring treatments	M^1	$\sum N_{i.}(Y_{i.})^2$
Treatments eliminating blocks	$N - 1$	$\sum_j Q_j \tilde{\alpha}_j$
Error	$N^* - M^1 - N + 1$	By subtraction
Total	N^*	$\sum_{ij}(Y_{ij})^2$

Hence, we have v as defined in (2.1) above is distributed as $F'_{p,q,\lambda}$, where p, q, and λ are as defined in (2.2) above.

Now it is clear that $\lambda = 0$ if and only if $\alpha_1 = \alpha_2 = \cdots = \alpha_N$ since λ is a sum of non-negative terms and can be zero if and only if each term of the sum is zero. Therefore to test the hypothesis $\alpha_1 = \alpha_2 = \cdots = \alpha_N$ we use v as Snedecor's F with p degrees of freedom and q degrees of freedom, where p and q are as defined in (2.2).

3. The analysis of variance. In Section 2 it was shown that the test function v could be used to test the hypothesis $\alpha_1 = \alpha_2 = \cdots = \alpha_N$. We will now show that v can be derived by the method of least squares. The model can be considered as a two-way classification model with unequal numbers in the subclasses. In this case the conventional analysis is given in Table 3.1 [4]. If we now denote the mean square for treatments eliminating blocks by T and the mean square for error by E, then $W = T/E$ is the least squares test function used to test the hypothesis $\alpha_1 = \alpha_2 = \cdots = \alpha_N$. ($N_{i.}$ is the number of treatments in the ith block). We will now show that the function v in Section 2 is the test criterion given by least squares.

In the above table, the Total SS minus the Block ignoring treatments SS is equal to

$$\sum_{t=1}^{k} q_t^6 - \sum_{t=1}^{k} q_t^5 .$$

It remains only to show that

$$\sum_{j=1}^{N} Q_j \tilde{\alpha}_j = \sum_{t=1}^{k} q_t^3 + \sum_{t=1}^{k-1} q_t^4$$

and the rest follows by subtraction ($\tilde{\alpha}_j$ is the least squares estimate of α_j).

We have the following system of normal equations:

(3.1.1) $\quad Y_{i.} = \tilde{\mu} + \tilde{\beta}_i + \tilde{\alpha}_.'^1, \qquad i = M^2 + 1, M^2 + 2, \cdots, M^1$

(3.1.2) $\quad Y_{i.} = \tilde{\mu} + \tilde{\beta}_i + \tilde{\alpha}_.'^2, \qquad i = M^3 + 1, M^3 + 2, \cdots, M^2$

$\vdots$

(3.1.$k - 1$) $\quad Y_{i.} = \tilde{\mu} + \tilde{\beta}_i + \tilde{\alpha}_.'^{k-1}, \qquad i = M^k + 1, M^k + 2, \cdots, M^{k-1}$

(3.1.k) $\quad Y_{i.} = \tilde{\mu} + \tilde{\beta}_i + \tilde{\alpha}_. , \qquad i = 1, 2, \cdots, M^k$

(3.2.1) $\quad Y_{.j}^1 = \tilde{\mu} + \tilde{\beta}_.^1 + \tilde{\alpha}_j^1 , \qquad j = 1, 2, \cdots, N^1$

(3.2.2) $\quad Y^2_{.j} = \hat{\mu} + \hat{\beta}^2_{.} + \hat{\alpha}^2_j, \qquad j = N^1 + 1, N^1 + 2, \cdots, N^2$

$\vdots$

(3.2.$k-1$) $\quad Y^{k-1}_{.j} = \hat{\mu} + \hat{\beta}^{k-1}_{.} + \hat{\alpha}^{k-1}_j, \qquad j = N^{k-2} + 1, N^{k-2} + 2, \cdots, N^{k-1}$

(3.2.k) $\quad Y^k_{.j} = \hat{\mu} + \hat{\beta}^k_{.} + \hat{\alpha}^k_j, \qquad j = N^{k-1} + 1, N^{k-1} + 2, \cdots, N^k$

where

$$\hat{\beta}^t_{.} = \frac{\sum_{i=1}^{M^t} \hat{\beta}_i}{M^t}.$$

Imposing the linear restriction $\hat{\alpha}_. = 0$, we find from (3.1.k) that

$$\hat{\mu} + \hat{\beta}_i = Y_{i.}, \qquad i = 1, 2, \cdots, M^k.$$

Substituting this into (3.2.k) we have

$$\hat{\alpha}^k_j = Y^k_{.j} - \frac{N^{k-1}Y'^{k-1}_{..} + N_k Y^k_{..}}{N}, \qquad j = N^{k-1} + 1, N^{k-1} + 2, \cdots, N^k$$

Now since

$$\sum_{j=1}^{N^{k-1}} \hat{\alpha}_j = - \sum_{j=N^{k-1}+1}^{N} \hat{\alpha}_j$$

under the restriction, $\hat{\alpha}_. = 0$, we may now substitute back and solve (3.1.$k-1$) obtaining

$$\hat{\mu} + \hat{\beta}_i = Y_{i.} + \frac{N_k}{N^{k-1}} \cdot \frac{N^{k-1}}{N} (Y^k_{..} - Y'^{k-1}_{..})$$

$$= Y_{i.} + \frac{N_k}{N} (Y^k_{..} - Y'^{k-1}_{..}), \qquad i = M^k + 1, M^k + 2, \cdots, M^{k-1}$$

Substituting back into (3.2.$k-1$) we get

$$\hat{\alpha}^{k-1}_j = Y^{k-1}_{.j} - \frac{M^k}{M^{k-1}} \left[\frac{N^{k-1}Y'^{k-1}_{..} + N_k Y^k_{..}}{N} \right] - \frac{\sum_{i=M^k+1}^{M^{k-1}} Y_{i.}}{M^{k-1}}$$

$$- \frac{M^{k-1} - M^k}{M^{k-1}} \cdot \frac{N_k}{N} (Y^k_{..} - Y'^{k-1}_{..})$$

$$= Y^{k-1}_{.j} - \frac{M^k}{M^{k-1}} Y'^{k-1}_{..} - \frac{\sum_{i=M^k+1}^{M^{k-1}} Y_{i.}}{M^{k-1}} - \frac{N_k}{N} (Y^k_{..} - Y'^{k-1}_{..})$$

$$= Y^{k-1}_{.j} - \frac{N^{k-2} Y'^{k-2}_{..} + N_{k-1} Y^{k-1}_{..}}{N^{k-1}} - \frac{N_k}{N} (Y^k_{..} - Y'^{k-1}_{..}),$$

$$j = N^{k-2} + 1, N^{k-2} + 2, \cdots, N^{k-1}.$$

Finishing the solution in this manner, we obtain

$$\hat{\alpha}^p_j = Y^p_{.j} - \frac{N^{p-1} Y'^{p-1}_{..} + N_p Y^p_{..}}{N^p} - \sum_{t=p+1}^{k} \frac{N_t}{N_t} (Y^t_{..} - Y'^{t-1}_{..}),$$

$$j = N^{p-1} + 1, N^{p-1} + 2, \cdots, N^p, \; p = 1, 2, \cdots, k.$$

which may be written as

$$\hat{a}_j^p = Y_{.j}^p - Y_{..}^p - \frac{N^{p-1}}{N^p}(Y_{..}'^{p-1} - Y_{..}^p) - \sum_{t=p+1}^{k} \frac{N_t}{N_t}(Y_{..}^t - Y_{..}'^{t-1}), \tag{3.3}$$

$$j = N^{p-1}+1, N^{p-1}+2, \cdots, N^p, p = 1, 2, \cdots, k$$

Now

$$Q_j^p = M^p Y_{.j}^p - \sum_{i=1}^{M^p} Y_{ij},$$

$$j = N^{p-1}+1, N^{p-1}+2, \cdots, N^p, p = 1, 2, \cdots, k$$

But

$$\frac{\sum_{i=1}^{M^p} Y_i}{M^p} = \frac{M^1}{M^p} \cdot \frac{N^{1-1} Y_{..}'^{1-1} + N_1 Y_{..}^1}{N^1} + \frac{\sum_{i=M^1+1}^{M^p} Y_i}{M^p},$$

Hence

$$Q_j^p = M^p \Big[Y_{.j}^p - Y^p - \frac{N^{p-1}}{N^p}(Y_{..}'^{p-1} - Y_{..}^p) - \sum_{t=p+1}^{k} \frac{M^t N_t}{M^p N^t}(Y_{..}^t - Y_{..}'^{t-1}) \Big], \quad p = 1, 2, \cdots, k.$$

Therefore

$$\sum_{j=1}^{N} Q_j \hat{a}_j = \sum_{p=1}^{k} \sum_{j=N^{p-1}+1}^{N^p} Q_j^p \hat{a}_j^p = \sum_{p=1}^{k} \Big[M^p \sum_{j=N^{p-1}+1}^{N^p} (Y_{.j}^p - Y_{..}^p)^2 \Big]$$

$$+ \sum_{p=1}^{k} M^p N_p \Big[\frac{N^{p-1}}{N^p}(Y_{..}'^{p-1} - Y_{..}^p) + \sum_{t=p+1}^{k} \frac{M^t N_t}{M^p N^t}(Y_{..}^t - Y_{..}'^{t-1}) \Big] \cdot \Big[\frac{N^{p-1}}{N^p}(Y_{..}'^{p-1} - Y_{..}^p) + \sum_{t=p+1}^{k} \frac{N_t}{N^t}(Y_{..}^t - Y_{..}'^{t-1}) \Big].$$

Collecting coefficients of $(Y_{..}'^r - Y_{..}^{r+1})^2$, we have

$$\frac{M^{r+1} N_{r+1} (N^r)^2}{(N^{r+1})^2} + \frac{M^r N_r M^{r+1} (N_{r+1})^2}{M^r (N^{r+1})^2} + \cdots + \frac{M^1 N_1 M^{r+1} (N_{r+1})^2}{M^1 (N^{r+1})^2}.$$

Combining the last r terms this becomes

$$\frac{M^{r+1} N_{r+1} (N^r)^2}{(N^{r+1})^2} + \frac{N^r M^{r+1} (N_{r+1})^2}{(N^{r+1})^2} = \frac{M^{r+1} N_{r+1} N^r (N^r + N_{r+1})}{(N^{r+1})^2}$$

$$= \frac{M^{r+1} N_{r+1} N^r}{N^{r+1}}$$

Collecting coefficients of $(Y'^{r}_{\cdot} - Y^{r+1}_{\cdot})(Y'^{s}_{\cdot} - Y^{s+1}_{\cdot})$, $r < s$, we have

$$-\frac{M^{r+1}N_{r+1}N^{r}N_{s+1}}{N^{r+1}N^{s+1}} + \frac{M^{r}N_{r}M^{r+1}N_{r+1}N_{s+1}}{M^{r}N^{r+1}N^{s+1}} + \cdots + \frac{M^{1}N_{1}M^{r+1}N_{r+1}N_{s+1}}{M^{1}N^{r+1}N^{s+1}}$$

$$-\frac{M^{r+1}N_{r+1}N^{r}M^{s+1}N_{s+1}}{N^{r+1}M^{r+1}N^{s+1}} + \frac{M^{r}N_{r}N_{r+1}M^{s+1}N_{s+1}}{N^{r+1}M^{r}N^{s+1}} + \cdots$$

$$+ \frac{M^{1}N_{1}N_{r+1}M^{s+1}N_{s+1}}{N^{r+1}M^{1}N^{s+1}}.$$

Combining all but the first term of each part gives

$$-\frac{M^{r+1}N_{r+1}N^{r}N_{s+1}}{N^{r+1}N^{s+1}} + \frac{N^{r}M^{r+1}N_{r+1}N_{s+1}}{N^{r+1}N^{s+1}} - \frac{N_{r+1}N^{r}M^{s+1}N_{s+1}}{N^{r+1}N^{s+1}}$$

$$+ \frac{N^{r}N_{r+1}M^{s+1}N_{s+1}}{N^{r+1}N^{s+1}} = 0.$$

Now since these two general terms are the only possible ones involved in the second summation of the expression for

$$\sum_{j=1}^{N} Q_j \bar{\alpha}_j,$$

we have

$$\sum_{j=1}^{N} Q_j \bar{\alpha}_j = \sum_{p=1}^{k} \left[M^{p} \sum_{j=N^{p-1}+1}^{N^{p}} (Y_j^{p} - Y_{\cdot}^{p})^2 \right]$$

$$+ \sum_{p=1}^{k-1} \frac{M^{p+1}N_{p+1}N^{p}}{N^{p+1}} (Y'^{p} - Y_{\cdot}^{p+1})^2$$

$$= \sum_{i=1}^{k} q_i^3 + \sum_{i=1}^{k-1} q_i^4$$

since $N^0 = 0$.

Now by subtraction the Error S.S. must be

$$\sum_{i=1}^{k} q_i^1 + \sum_{i=1}^{k-1} q_i^2.$$

Also since the degrees of freedom for error and treatments eliminating blocks in Table 3.1 are the same as q and p of (2.2), then we have $W' = v$. Thus we have shown that the test function given in section 2 is that given by the method of least squares.

4. *Means and standard errors.* We will now derive the best, linear, unbiased estimates of $\alpha_s - \alpha_t$ and the standard errors of these estimates.

THEOREM II.

$$\bar{\alpha}_s^p = Y_{\cdot s}^p - Y_{\cdot}^p - \frac{N^{p-1}}{N^p}(Y'^{p-1}_{\cdot} - Y_{\cdot}^p) - \sum_{i=p+1}^{k} \frac{N_i}{N^i}(Y_{\cdot}^i - Y'^{i-1}_{\cdot}),$$

which may be written as

$$\bar{\alpha}_j^p = Y_{.j}^p - Y_{..}^p - \frac{N^{p-1}}{N^p}(Y_{..}'^{p-1} - Y_{..}^p) - \sum_{t=p+1}^{k} \frac{N_t}{N_t}(Y_{..}^t - Y_{..}'^{t-1}), \tag{3.3}$$

$$j = N^{p-1}+1, N^{p-1}+2, \cdots, N^p, p = 1, 2, \cdots, k$$

Now

$$Q_j^p = M^p Y_{.j}^p - \sum_{i=1}^{M^p} Y_{i.},$$

$$j = N^{p-1}+1, N^{p-1}+2, \cdots, N^p, p = 1, 2, \cdots, k$$

But

$$\frac{\sum_{i=1}^{M^p} Y_{i.}}{M^p} = \frac{M^k}{M^p} \cdot \frac{N^{k-1}\, Y_{..}'^{k-1} + N_k\, Y_{..}^k}{N^k} + \frac{\sum_{i=M^k+1}^{M^p} Y_{i.}}{M^p},$$

Hence

$$Q_j^p = M^p \Bigg[Y_{.j}^p - Y_{..}^p - \frac{N^{p-1}}{N^p}(Y_{..}'^{p-1} - Y_{..}^p)$$

$$- \sum_{t=p+1}^{k} \frac{M^t N_t}{M^p N^t}(Y_{..}^t - Y_{..}'^{t-1}) \Bigg], \quad p = 1, 2, \cdots, k.$$

Therefore

$$\sum_{j=1}^{N} Q_j\, \bar{\alpha}_j = \sum_{p=1}^{k} \sum_{j=N^{p-1}+1}^{N^p} Q_j^p\, \bar{\alpha}_j^p = \sum_{p=1}^{k} \Bigg[M^p \sum_{j=N^{p-1}+1}^{N^p} (Y_{.j}^p - Y_{..}^p)^2 \Bigg]$$

$$+ \sum_{p=1}^{k} M^p N_p \Bigg[\frac{N^{p-1}}{N^p}(Y_{..}'^{p-1} - Y_{..}^p)$$

$$+ \sum_{t=p+1}^{k} \frac{M^t N_t}{M^p N^t}(Y_{..}^t - Y_{..}'^{t-1}) \Bigg] \cdot \Bigg[\frac{N^{p-1}}{N^p}(Y_{..}'^{p-1} - Y_{..}^p)$$

$$+ \sum_{t=p+1}^{k} \frac{N_t}{N^t}(Y_{..}^t - Y_{..}'^{t-1}) \Bigg].$$

Collecting coefficients of $(Y_{..}'^r - Y_{..}^{r+1})^2$, we have

$$\frac{M^{r+1} N_{r+1} (N^r)^2}{(N^{r+1})^2} + \frac{M^r N_r M^{r+1} (N_{r+1})^2}{M^r (N^{r+1})^2} + \cdots + \frac{M^1 N_1 M^{r+1} (N_{r+1})^2}{M^1 (N^{r+1})^2}.$$

Combining the last r terms this becomes

$$\frac{M^{r+1} N_{r+1} (N^r)^2}{(N^{r+1})^2} + \frac{N^r M^{r+1} (N_{r+1})^2}{(N^{r+1})^2} = \frac{M^{r+1} N_{r+1} N^r (N^r + N_{r+1})}{(N^{r+1})^2}$$

$$= \frac{M^{r+1} N_{r+1} N^r}{N^{r+1}}$$

Collecting coefficients of $(Y'^{r}_{..} - Y^{r+1}_{..})(Y'^{s}_{..} - Y^{s+1}_{..})$, $r < s$, we have

$$-\frac{M^{r+1} N_{r+1} N^r N_{s+1}}{N^{r+1} N^{s+1}} + \frac{M^r N_r M^{r+1} N_{r+1} N_{s+1}}{M^r N^{r+1} N^{s+1}} + \cdots + \frac{M^1 N_1 M^{r+1} N_{r+1} N_{s+1}}{M^1 N^{r+1} N^{s+1}}$$
$$-\frac{M^{r+1} N_{r+1} N^r M^{s+1} N_{s+1}}{N^{r+1} M^{r+1} N^{s+1}} + \frac{M^r N_r N_{r+1} M^{s+1} N_{s+1}}{N^{r+1} M^r N^{s+1}} + \cdots$$
$$+ \frac{M^1 N_1 N_{r+1} M^{s+1} N_{s+1}}{N^{r+1} M^1 N^{s+1}}.$$

Combining all but the first term of each part gives

$$-\frac{M^{r+1} N_{r+1} N^r N_{s+1}}{N^{r+1} N^{s+1}} + \frac{N^r M^{r+1} N_{r+1} N_{s+1}}{N^{r+1} N^{s+1}} - \frac{N_{r+1} N^r M^{s+1} N_{s+1}}{N^{r+1} N^{s+1}}$$
$$+ \frac{N^r N_{r+1} M^{s+1} N_{s+1}}{N^{r+1} N^{s+1}} = 0.$$

Now since these two general terms are the only possible ones involved in the second summation of the expression for

$$\sum_{j=1}^{N} Q_j \bar{\alpha}_j ,$$

we have

$$\sum_{j=1}^{N} Q_j \bar{\alpha}_j = \sum_{p=1}^{k} \left[M^p \sum_{j=N^{p-1}+1}^{N^p} (Y^p_j - Y^p_{..})^2 \right]$$
$$+ \sum_{p=1}^{k-1} \frac{M^{p+1} N_{p+1} N^p}{N^{p+1}} (Y'^{p}_{..} - Y^{p+1}_{..})^2$$
$$= \sum_{i=1}^{k} q_i^3 + \sum_{i=1}^{k-1} q_i^4$$

since $N^0 = 0$.

Now by subtraction the Error S.S. must be

$$\sum_{i=1}^{k} q_i^1 + \sum_{i=1}^{k-1} q_i^2 .$$

Also since the degrees of freedom for error and treatments eliminating blocks in Table 3.1 are the same as q and p of (2.2), then we have $W = v$. Thus we have shown that the test function given in section 2 is that given by the method of least squares.

4. Means and standard errors. We will now derive the best, linear, unbiased estimates of $\alpha_s - \alpha_t$ and the standard errors of these estimates.

THEOREM II.

$$\bar{\alpha}^p_s = Y^p_{.s} - Y^p_{..} - \frac{N^{p-1}}{N^p} (Y'^{p-1}_{..} - Y^p_{..}) - \sum_{t=p+1}^{k} \frac{N_t}{N^t} (Y^t_{..} - Y'^{t-1}_{..}),$$

$$s = N^{p-1} + 1, N^{p-1} + 2, \cdots, N^p$$
$$p = 1, 2, \cdots, k$$

is the best, linear, unbiased estimate of $\alpha_s - \alpha_.$, *and therefore* $\bar{\alpha}_s - \bar{\alpha}_u$ *is the best, linear, unbiased estimate of* $\alpha_s - \alpha_u$.

PROOF. Since $\bar{\alpha}_s$ was found by the method of least squares (3.3) using the linear restriction $\bar{\alpha}_. = 0$, this result is a consequence of the Markoff Theorem [3].

THEOREM III. *The variance of the estimate of* $\alpha_s^p - \alpha_u^p$ *is* $2\sigma^2/M^p$ *if* $s \neq u$, *and* $s, u = N^{p-1} + 1, N^{p-1} + 2, \cdots, N^p$ *for* $p = 1, 2, \cdots, k$. *The variance of the estimate of* $\alpha_s^p - \alpha_u^r$ *is*

$$\sigma^2 \left[\frac{N^p - 1}{M^p N^p} + \frac{N^{r-1} + 1}{M^r N^{r-1}} + \sum_{t=p+1}^{r-1} \frac{N_t}{M^t N^t N^{t-1}} \right]$$

for $s = N^{p-1} + 1, N^{p-1} + 2, \cdots, N^p$, $u = N^{r-1} + 1, N^{r-1} + 2, \cdots, N^r$, $p = 1, 2, \cdots, k - 1$, $r = p + 1, p + 2, \cdots, k$.

PROOF.

$$\bar{\alpha}_s^p - \bar{\alpha}_u^p = Y_{.s}^p - Y_{.u}^p,$$

$s \neq t$ and $s, u = N^{p-1} + 1, N^{p-1} + 2, \cdots, N^p$, $p = 1, 2, \cdots, k$.

And

$$\mathrm{Var}\,(\bar{\alpha}_s^p - \bar{\alpha}_u^p) = E\left[\frac{\sum_{i=1}^{M^p} c_{is}}{M^p} \frac{\sum_{i=1}^{M^p} c_{iu}}{M^p}\right]$$
$$= \frac{\sigma^2}{M^p} + \frac{\sigma^2}{M^p} = \frac{2\sigma^2}{M^p}, \qquad p = 1, 2, \cdots, k.$$

Now

$$\bar{\alpha}_s^p - \bar{\alpha}_u^r = Y_{.s}^p - \frac{N^{p-1} Y_{..}^{\prime p-1} + N_p Y_{..}^p}{N^p} - \sum_{t=p+1}^{r-1} \frac{N_t}{N^t} (Y_{..}^t - Y_{..}^{\prime t-1}) - (Y_{.u}^r + Y_{..}^{\prime r-1})$$

and by straightforward application of expected values we arrive at the result.

From the theory of least squares it follows that the error mean square

$$\frac{1}{q} \left| \sum_{t=1}^{k} q_t^1 + \sum_{t=1}^{k-1} q_t^2 \right|$$

(where q is defined in 2.2) is an unbiased estimate of σ^2 and is independent of $\bar{\alpha}_i$. Therefore, these quantities may be used to set confidence limits about the difference between treatment means or any linear contrast of treatment means. Therefore, by using equation (2.1), the analysis of variance for the Staircase Design is easily computed. By using the formulas in Theorem II and Theorem III, the means and standard errors can be easily computed even if the number of steps is large. In another paper we will give detailed computing instructions with a numerical example of the Staircase Design.

the best, linear, unbiased estimate o
ıear, unbiased estimate of $\alpha_s - \alpha_u$.
PROOF. Since $\tilde{\alpha}_s$ was found by t
ıear restriction $\tilde{\alpha}_. = 0$, this resul
].

THEOREM III. *The variance of th*
ıd $s, u = N^{p-1} + 1, N^{p-1} + 2,$
the estimate of $\alpha_s^p - \alpha_u^r$ *is*

$$\sigma^2 \left[\frac{N^p - 1}{M^p N^p} + \right.$$

r $s = N^{p-1} + 1, N^{p-1} + 2, \cdots$
$2, \cdots, k - 1, r = p + 1, p$
PROOF.

$$\tilde{\alpha}_s^p -$$

$\neq t$ and $s, u = N^{p-1} + 1, N$

ıd

$$\text{Var}(\tilde{\alpha}_s^p - \tilde{\alpha}$$

ow

$$- \tilde{\alpha}_u^r = Y_{.s}^p - \frac{N^{p-1} Y'^{p-1}_{..}}{N}$$

ıd by straightforward appli
From the theory of least s

here q is defined in 2.2) i
. Therefore, these quanti
fference between treatmen
ıerefore, by using equati
esign is easily computed.
I, the means and standar
steps is large. In anothe
ith a numerical example e

A NOTE ON THE FUNDAMENTAL IDENTITY OF SEQUENTIAL ANALYSIS

By R. R. Bahadur

University of Chicago[1]

1. Introduction. This note points out that the fundamental identity of sequential analysis [1] can be regarded as a special case of a formula for the probability that the sampling terminates at some finite stage. This viewpoint, explored in Sections 2 and 3, provides proofs of the identity, and of its differentiability under the expectation sign, that seem more intuitive than the proofs in the literature ([1], [2], [3], [5], [6]).

The formula also has application to the well-known problem (cf., e.g., [7], [8]) of evaluating the probability of eventual termination of a random walk on the real line, in the case when there is one fixed barrier and a drift away from the barrier. Some upper and lower bounds on the probability in question are obtained in Section 4.

In concluding this introduction, the writer wishes to thank his colleague L. J. Savage for discussions and suggestions that have made a substantial contribution to this work.

2. A Formula for $P(n < \infty)$. Let x be a real valued random variable with distribution function F. It is assumed that the moment generating function

$$\phi(t) = \int_{-\infty}^{\infty} e^{tx}\, dF \tag{1}$$

exists for every real t in some neighbourhood of $t = 0$. Throughout this note, t is restricted to real values for which ϕ exists.

Let $x_{(\infty)} = (x_1, x_2, \cdots, \text{ad inf})$ denote a sequence of independent and identically distributed observations on x. Consider a fixed sequential sampling procedure, that is, a set of rules for observing the components $x_1, x_2, \cdots$, of $x_{(\infty)}$ one by one, such that at each stage the decision whether experimentation is to continue is a (possibly randomised) function of the observed values in hand at that stage. (Cf., e.g., [6], [9]). Let n denote the total number of components x_m observed in a given instance. It is assumed that the sampling procedure is closed under F, that is,

$$P(n < \infty \mid F) = 1. \tag{2}$$

The procedure is otherwise arbitrary.

Write $s = x_1 + \cdots + x_n$ and

$$\psi(t, n, s) = [\phi(t)]^{-n} e^{ts} \tag{3}$$

if $n < \infty$, and write $\psi = 1$ (say) if $n = \infty$.

Received February 7, 1957; revised July 12, 1957.

[1] Now at the Indian Statistical Institute, Calcutta.

THEOREM 1. *For every t,*

$$E(\psi(t, n, s) \mid F) = P(n < \infty \mid G_t), \tag{4}$$

where

$$dG_t = [\phi(t)]^{-1} e^{ts}\, dF. \tag{5}$$

For each t, G_t defined by (5) is clearly a probability distribution function, so that $\{G_t\}$ is an exponential family of alternative distributions of x, with F a member of this family. Such families of distributions have been studied in various statistical contexts, including that of sequential analysis (cf., e.g., [1], [10], [11]). It may be added, however, that the notion of alternative distributions is not essential to this paper, and the introduction here of the family $\{G_t\}$ could be regarded as a device in the study of the given sampling rule when F obtains. This device is, of course, a familiar one in probability theory (cf., e.g., [3], [12], [13], [14]).

To establish Theorem 1, for each $m = 1, 2, \cdots$ let $R^{(m)}$ denote the nonsequential sample space of exactly m observations, that is, of points

$$(x_1, \cdots, x_m) = x_{(m)}$$

say. For each m, let $\alpha_m(x_{(m)})$ be the conditional probability of the event $n = m$ given $x_{(\infty)}$. The sequence $\alpha_1, \alpha_2, \cdots$ of functions on $R^{(1)}, R^{(2)}, \cdots$ characterizes the given sampling procedure, and is, of course, independent of the distribution of x (cf., e.g., [9]). For each m, let $F^{(m)}$ denote the distribution function of $x_{(m)}$ when F obtains, that is, $F^{(m)}(x_1, \cdots, x_m) = \prod_{i=1}^{m} F(x_i)$.

Let v denote the total outcome of the sequential experiment, that is, $v = (x_1, \cdots, x_n)$ if $n < \infty$ and $v = x_{(\infty)}$ if $n = \infty$. We note that if h is a real valued function of v such that $E(|h| \mid F) < \infty$, and (2) holds, then

$$E(h \mid F) = \sum_{m=1}^{\infty} \int_{R^{(m)}} \alpha_m \cdot h_m \, dF^{(m)} \tag{6}$$

where $h_1, h_2, \cdots$ is the (essentially unique) sequence of functions on $R^{(1)}, R^{(2)}, \cdots$ such that $h_m = h$ when $n = m$, (cf., e.g., [9]). The right side of (6) is an absolutely convergent series; in fact, h is integrable if and only if

$$\sum_m \int \alpha_m \cdot |h_m| \, dF^{(m)} < \infty.$$

In accordance with the above notation, for each $m = 1, 2, \cdots$ let s_m denote the nonsequential random variable $x_1 + \cdots + x_m$. Then it is easy to establish (4) thus:

$$\begin{aligned} P(n < \infty \mid G_t) &= \sum_{m=1}^{\infty} P(n = m \mid G_t) \\ &= \sum_{m=1}^{\infty} \int_{R^{(m)}} \alpha_m \cdot dG_t^{(m)} \\ &= \sum_{m=1}^{\infty} \int_{R^{(m)}} \alpha_m \cdot \phi^{-m} \cdot e^{ts_m} \, dF^{(m)} \quad \text{by (5)} \\ &= E(\psi \mid F) \quad \text{by (2), (3), and (6).} \end{aligned} \tag{7}$$

3. Wald's identity. It follows from Theorem 1 that Wald's identity, namely

$$E[\psi(t, n, s) \mid F] = 1, \tag{8}$$

holds for a given value of t if and only if

$$P(n < \infty \mid G_t) = 1 \tag{9}$$

for the same t value.

It follows easily from the preceding remark that a sufficient condition for the validity of (8) for all t is that there exist a finite k such that $P(n < k \mid F) = 1$. The same remark, together with the strong law of large numbers and the law of the iterated logarithm (cf. e.g., [15]), also yields the following sufficient condition for the same conclusion (assuming that F is a non-degenerate distribution): there exists an h, $-\infty < h < \infty$, and two sequences $\{a_m\}$ and $\{b_m\}$ such that

(i) $a_m = mh + o(\sqrt{m \log \log m})$, $b_m = mh + o(\sqrt{m \log \log m})$ as $m \to \infty$ and

(ii) for each $x_{(\infty)} = (x_1, x_2, \cdots)$, either $n < \infty$ or $a_m < s_m < b_m$ for all sufficiently large m. This condition seems weaker than other structural conditions of the same type in the literature for the validity of (8) for all t.

It may be noted here that the first paragraph of this section also suggests examples where Wald's identity fails to hold for all t. This was pointed out to the writer by L. J. Savage. The discussion in Section 4 concerns a general example of this sort.

Next, we shall describe an alternative sufficient condition for the validity of (9) (and thereby of (8)) in an assigned neighbourhood of $t = 0$, given (2) and (5). This condition does not depend on the detailed structure of the sampling rule; it is, essentially, that under F the joint moment generating function of n and s exist in a sufficiently large neighbourhood of the origin. As it happens, the condition also assures the validity of differentiation under the expectation sign in (8), that is, $D^k\psi(t, n, s)$ is integrable and

$$E(D^k\psi \mid F) = 0 \quad \text{for} \quad k = 1, 2, \cdots \tag{10}$$

and each t in the neighbourhood, where $D^k \equiv d^k / dt^k$.

Let I be an open interval including $t = 0$ such that $\phi(t)$ exists for each t in I.

THEOREM 2. *Suppose that corresponding to each t in I there exists a*

$$z > -\log_e \phi(t) \tag{11}$$

such that

$$E(e^{ts+zn} \mid F) < \infty. \tag{12}$$

Then (8), (9), *and* (10) *hold for t in I.*

The proof of Theorem 2 will be indicated later, but first some remarks by way of discussion of its hypothesis.

REMARK 1. Let C denote the set of all points (t, z) in the plane such that

(12) holds. Then (a) C is a convex set, (b) $(t, z) \varepsilon C$ implies $(t, z_0) \varepsilon C$ for all $z_0 \leqq z$, since n is non-negative, and (c) each point on the graph of the function $z = -\log \phi(t)$ is in C, by (3) and (4). It follows, in particular, that (d) the *hypothesis of Theorem 2 is that the graph of* $-\log \phi(t)$ *lie in the interior of* C, at least when t is restricted to I. It should be noted also (e) that $z = -\log \phi$ is a concave function of t, possessing derivatives of all orders, with $z(0) = 0$, and $z'(0) = -E(x \mid F)$, by (1). The facts (a), (b), (c), (d), and (e) are useful in the proof of Theorem 2, and also in the deduction of special sufficient conditions for the validity of (9) and (10). (Cf. remarks 3 and 4 below).

REMARK 2. In the statement of Theorem 2, as also in remark 1(d) above, the hypothesis is stated in terms of the given distribution F. The hypothesis can also be stated in terms of the associated distributions G_t, as follows: for each t in I, the *conditional* moment generating function of n given $n < \infty$ exists in some neighbourhood of zero when G_t obtains, that is, $E(e^{\delta n} \mid G_t, n < \infty) = \sum_m e^{\delta m} P(n = m \mid G_t, n < \infty) < \infty$ for some $\delta > 0$. This alternative formulation follows readily from (5), (6), (11) and (12).

REMARK 3. Suppose that ϕ exists for all t. If

$$E(e^{zn} \mid F) < \infty \tag{13}$$

for some $z > \sup \{-\log \phi(t): -\infty < t < \infty\}$, and if

$$E(e^{ts} \mid F) < \infty \tag{14}$$

for all t, then (8), (9), and (10) hold for all t. A stronger sufficient condition for the same conclusion is that (13) hold for all z

REMARK 4. If (13) holds for some $z > 0$, then n and s possess moments of all orders, and there exists a neighbourhood of $t = 0$ in which (8), (9) and (10) hold. A stronger sufficient condition for the same conclusion is that $E(x \mid F) \neq 0$ and (14) holds for some t of the same sign as $E(x \mid F)$ These conditions are of interest since the validity of (10) in a neighbourhood of $t = 0$ is sufficient for most applications (cf. [2]) of differentiation

REMARK 5. A theorem of Albert [3], [4] states that if ϕ exists for all t, if

$$P(x > 0 \mid F) > 0$$

and $P(x < 0 \mid F) > 0$, and if the sampling procedure is a random walk based on the cumulative sums s_m (with fixed barriers a and b, $a < 0 < b$), then (8) and (10) hold for all t. It can be shown (cf. Lemma 2 of [3] and Remark 3 above) that in this case the hypothesis of Theorem 2 is satisfied by $I = (-\infty, \infty)$, so that Albert's theorem is a special case of Theorem 2.

REMARK 6. In his applications of martingale theory to sequential analysis, Doob [6] derives from Theorem 2.2 of [6] a sufficient condition for the validity of (8) for a given t. It can be shown that if this condition holds for each t in I then the hypothesis of Theorem 2 is satisfied A fuller discussion of the relation between Doob's Theorem 2.2 in its application to the present case and Theorems 1 and 2 of this note would be worthwhile, but cannot be undertaken here.

REMARK 7. It is easy to see that $P(n < k \mid F) = 1$ for some finite k implies the hypothesis of Theorem 2. L. J. Savage has constructed examples showing that the other condition stated in the second paragraph of this section does not imply the hypothesis of Theorem 2; in fact, (10) fails for $k = 2$ in these examples.

We turn now to the proof of Theorem 2. The first step is to show that (9) holds in a sufficiently small neighbourhood of zero, that is,

$$\sum_{m=1}^{\infty} p_m(t) = 1 \tag{15}$$

for all t in the neighbourhood, where $p_m(t)$ is an abbreviation of $P(n = m \mid G_t)$, so that

$$p_m(t) = \int_{R^{(m)}} \alpha_m \cdot \{\phi(t)\}^{-m} \cdot e^{ts_m}\, dF^{(m)} \tag{16}$$

by (5). Write $\beta_m = 1 - \sum_{i=1}^{m} \alpha_i$, and $\rho(t) = \phi(2t)/\phi^2(t)$. Then, for any t in I and any m,

$$\begin{aligned} P(n > m \mid G_t) &= \int_{R^{(m)}} \beta_m \, dG_t^{(m)} \quad \text{since} \quad \beta_m = P(n > m \mid x_{(m)}) \\ &= \phi^{-m} \int_{R^{(m)}} \beta_m e^{ts_m}\, dF^{(m)} \quad \text{by} \quad (5) \\ &\leqq \phi^{-m} \left\{ \int_{R^{(m)}} \beta_m^2 \, dF^{(m)} \right\}^{1/2} \cdot \left\{ \int_{R^{(m)}} e^{2ts_m} \, dF^{(m)} \right\}^{\frac{1}{2}} \\ &= \rho^{m/2} \left\{ \int_{R^{(m)}} \beta_m^2 \, dF^{(m)} \right\}^{1/2} \\ &\leqq \rho^{m/2} \left\{ \int_{R^{(m)}} \beta_m \, dF^{(m)} \right\}^{1/2} \quad \text{since} \quad 0 \leqq \beta_m \leqq 1 \\ &= \sqrt{\rho(t)^m \cdot P(n > m \mid F)}. \end{aligned} \tag{17}$$

It follows from (11) and (12) with $t = 0$ that, for some $\lambda > 1$,

$$\lambda^m P(n > m \mid F) \to 0$$

as $m \to \infty$. Hence, by (17), $P(n > m \mid G_t) \to 0$ as $m \to \infty$ for each t such that $\rho(t) \leqq \lambda$. Thus (15) holds whenever $\rho(t) \leqq \lambda$. This establishes the desired conclusion, since $\lambda > 1$, ρ is continuous, and $\rho(0) = 1$.

The next step is to extend the validity of (15) to all t in I by analytic continuation, as follows. Let w denote the complex variable $t + iu$, and let $\phi(w)$ be defined by (1) in the strip $\{w : t \in I\}$. For each m let $p_m(w)$ be defined by (16) whenever $\phi(w)$ is defined and $\neq 0$. It follows from the differentiability of moment

generating functions that the functions p_m are differentiable everywhere in their domain of definition. A straightforward argument based on the continuity of ϕ, Remark 1(d) above, formula (6), and the convexity of the exponential function shows that corresponding to each t in I there exists a complex neighbourhood of t, $N(t)$ say, and a convergent series of positive terms, $\sum_m c_m(t)$ say, such that $|p_m(w)| \leqq c_m(t)$ for all $w \varepsilon N(t)$ and each $m = 1, 2, \cdots$. The details of this argument are omitted. It follows hence that $\sum_m p_m(w)$ is a uniformly convergent series of analytic functions, so that $\sum_m p_m(w)$ is well defined and analytic in $N(t)$. (Cf. e.g., [16]). Since this holds for each t, it follows from the preceding paragraph that $\sum_m p_m(w) = 1$ for $w \varepsilon N(t)$ and $t \varepsilon I$; in particular, (15) holds for all $t \varepsilon I$.

It follows from uniform convergence (cf. [16]) and the conclusion of the preceding paragraph that for each t in I and every $k = 1, 2, \cdots$,

$$\sum_m (d^k / dw^k) p_m(w)$$

is well defined and $= 0$ for $w \varepsilon N(t)$; in particular, $\sum_m D^k p_m(t) = 0$. Since, as is readily seen, D^k commutes with the integral sign in (16), we have

$$\sum_m D^k p_m(t) = \sum_m \int_{R^{(m)}} \alpha_m \cdot D^k \{\phi^{-m} e^{t s_m}\} \, dF^{(m)} \tag{18}$$
$$= 0 \text{ for } k = 1, 2, \cdots$$

and each t in I. Assuming for the moment that each of the functions $D^k\psi(t, n, s)$ is integrable when F obtains, it follows by inspection from (3), (6), and (18) that (10) holds for each t in I.

The next and final step in the proof is therefore to verify that each

$$D^k\psi(t, n, s)$$

is integrable. Since $D^k\psi$ is of the form $\psi \cdot \eta$ where η is a polynomial in n and s, it suffices to show that, for each t, $\psi \cdot | s |^i \cdot n^j$ is integrable for $i, j = 0, 1, 2, \cdots$. This may be established by showing that corresponding to each t in I there exist positive numbers ϵ and δ such that $\psi(t, n, s) \cdot \exp(\epsilon | s | + \delta n)$ is integrable. Since $\exp(\epsilon | s |) < \exp(\epsilon s) + \exp(-\epsilon s)$, it is easily seen from (3) and Remark 1(d) that this last condition is satisfied.

In concluding this section, we remark that Theorems 1 and 2 can be generalized, by straightforward extensions of the arguments used here, to the case when $x_1, x_2, \cdots$ is a sequence of independent but not necessarily identically distributed random variables, and each x_m takes values in k-dimensional Cartesian space, $1 \leqq k < \infty$. Another straightforward generalization that may be worth mentioning here is to the case where the sampling rule is defined for a sequence $w_1, w_2, \cdots$ of independent abstract random variables, and for each $i = 1, 2, \cdots x_i$ is a real (or vector) function of w_i.

4. An application. Let c be a positive constant, and let the sampling rule be defined thus: for any sequence $x_{(\infty)} = (x_1, x_2, \cdots)$, n is the smallest integer

REMARK 7. It is easy to see that $P(n < k \mid F) = 1$ for some finite k implies the hypothesis of Theorem 2. L. J. Savage has constructed examples showing that the other condition stated in the second paragraph of this section does not imply the hypothesis of Theorem 2; in fact, (10) fails for $k = 2$ in these examples.

We turn now to the proof of Theorem 2. The first step is to show that (9) holds in a sufficiently small neighbourhood of zero, that is,

$$\sum_{m=1}^{\infty} p_m(t) = 1 \tag{15}$$

for all t in the neighbourhood, where $p_m(t)$ is an abbreviation of $P(n = m \mid G_t)$, so that

$$p_m(t) = \int_{R^{(m)}} \alpha_m \cdot \{\phi(t)\}^{-m} \cdot e^{ts_m}\, dF^{(m)} \tag{16}$$

by (5). Write $\beta_m = 1 - \sum_{i=1}^{m} \alpha_i$, and $\rho(t) = \phi(2t)/\phi^2(t)$. Then, for any t in I and any m,

$$\begin{aligned} P(n > m \mid G_t) &= \int_{R^{(m)}} \beta_m\, dG_t^{(m)} \quad \text{since} \quad \beta_m = P(n > m \mid x_{(m)}) \\ &= \phi^{-m} \int_{R^{(m)}} \beta_m e^{ts_m}\, dF^{(m)} \quad \text{by} \quad (5) \\ &\leqq \phi^{-m} \left\{ \int_{R^{(m)}} \beta_m^2\, dF^{(m)} \right\}^{1/2} \cdot \left\{ \int_{R^{(m)}} e^{2ts_m}\, dF^{(m)} \right\}^{\frac{1}{2}} \\ &= \rho^{m/2} \left\{ \int_{R^{(m)}} \beta_m^2\, dF^{(m)} \right\}^{1/2} \\ &\leqq \rho^{m/2} \left\{ \int_{R^{(m)}} \beta_m\, dF^{(m)} \right\}^{1/2} \quad \text{since} \quad 0 \leqq \beta_m \leqq 1 \\ &= \sqrt{\rho(t)^m \cdot P(n > m \mid F)}. \end{aligned} \tag{17}$$

It follows from (11) and (12) with $t = 0$ that, for some $\lambda > 1$,

$$\lambda^m P(n > m \mid F) \to 0$$

as $m \to \infty$. Hence, by (17), $P(n > m \mid G_t) \to 0$ as $m \to \infty$ for each t such that $\rho(t) \leqq \lambda$. Thus (15) holds whenever $\rho(t) \leqq \lambda$. This establishes the desired conclusion, since $\lambda > 1$, ρ is continuous, and $\rho(0) = 1$.

The next step is to extend the validity of (15) to all t in I by analytic continuation, as follows. Let w denote the complex variable $t + iu$, and let $\phi(w)$ be defined by (1) in the strip $\{w : t \in I\}$. For each m let $p_m(w)$ be defined by (16) whenever $\phi(w)$ is defined and $\neq 0$. It follows from the differentiability of moment

generating functions that the functions p_m are differentiable everywhere in their domain of definition. A straightforward argument based on the continuity of ϕ, Remark 1(d) above, formula (6), and the convexity of the exponential function shows that corresponding to each t in I there exists a complex neighbourhood of t, $N(t)$ say, and a convergent series of positive terms, $\sum_m c_m(t)$ say, such that $|p_m(w)| \leqq c_m(t)$ for all $w \varepsilon N(t)$ and each $m = 1, 2, \cdots$. The details of this argument are omitted. It follows hence that $\sum_m p_m(w)$ is a uniformly convergent series of analytic functions, so that $\sum_m p_m(w)$ is well defined and analytic in $N(t)$. (Cf. e.g., [16]). Since this holds for each t, it follows from the preceding paragraph that $\sum_m p_m(w) = 1$ for $w \varepsilon N(t)$ and $t \varepsilon I$; in particular, (15) holds for all $t \varepsilon I$.

It follows from uniform convergence (cf. [16]) and the conclusion of the preceding paragraph that for each t in I and every $k = 1, 2, \cdots$,

$$\sum_m (d^k / dw^k) p_m(w)$$

is well defined and $= 0$ for $w \varepsilon N(t)$; in particular, $\sum_m D^k p_m(t) = 0$. Since, as is readily seen, D^k commutes with the integral sign in (16), we have

$$\sum_m D^k p_m(t) = \sum_m \int_{R^{(m)}} \alpha_m \cdot D^k \{\phi^{-m} e^{t s_m}\} \, dF^{(m)} \tag{18}$$
$$= 0 \text{ for } k = 1, 2, \cdots$$

and each t in I. Assuming for the moment that each of the functions $D^k \psi(t, n, s)$ is integrable when F obtains, it follows by inspection from (3), (6), and (18) that (10) holds for each t in I.

The next and final step in the proof is therefore to verify that each

$$D^k \psi(t, n, s)$$

is integrable. Since $D^k \psi$ is of the form $\psi \cdot \eta$ where η is a polynomial in n and s, it suffices to show that, for each t, $\psi \cdot |s|^i \cdot n^j$ is integrable for $i, j = 0, 1, 2, \cdots$. This may be established by showing that corresponding to each t in I there exist positive numbers ϵ and δ such that $\psi(t, n, s) \cdot \exp(\epsilon |s| + \delta n)$ is integrable. Since $\exp(\epsilon |s|) < \exp(\epsilon s) + \exp(-\epsilon s)$, it is easily seen from (3) and Remark 1(d) that this last condition is satisfied.

In concluding this section, we remark that Theorems 1 and 2 can be generalized, by straightforward extensions of the arguments used here, to the case when $x_1, x_2, \cdots$ is a sequence of independent but not necessarily identically distributed random variables, and each x_m takes values in k-dimensional Cartesian space, $1 \leqq k < \infty$. Another straightforward generalization that may be worth mentioning here is to the case where the sampling rule is defined for a sequence $w_1, w_2, \cdots$ of independent abstract random variables, and for each $i = 1, 2, \cdots x_i$ is a real (or vector) function of w_i.

4. An application. Let c be a positive constant, and let the sampling rule be defined thus: for any sequence $x_{(\infty)} = (x_1, x_2, \cdots)$, n is the smallest integer

m such that $s_m = x_1 + \cdots + x_m > c$, and $n = \infty$ if no such integer exists. Let G be a given distribution function such that $P(x > 0 \mid G) > 0$ and

$$P(x < 0 \mid G) > 0,$$

and such that $E(x \mid G)$ exists and is negative. It is shown in this section that if G admits a moment generating function, Theorem 1 can be used to obtain upper and lower bounds for $P(n < \infty \mid G)$. In certain special cases this method yields the exact value of $P(n < \infty \mid G)$.

The probability in question can be interpreted as the probability of ultimate ruin in playing an advantageous gamble long enough (with x_m the amount lost in the mth play, each x_m distributed according to G, and c the initial fortune of the player), and has been studied in connection with insurance theory (cf. [8], [12]). In [8] Dubourdieu has given derivations and original references to an upper bound, due to de Finetti, for this probability. The upper bounds obtained here are improvements of de Finetti's.

It is assumed henceforth that

$$\eta(h) = \int_{-\infty}^{\infty} e^{hx}\, dG \tag{19}$$

exists in a neighbourhood of $h = 0$. It then follows from the preceding hypotheses concerning G, by well-known properties of moment generating functions, that there exist uniquely determined points a and b (say), in the interior of the interval in which η exists, such that $0 < a < b$, $\eta'(a) = 0$, and $\eta(b) = 1$. We note that

$$\eta'(h) \geqq 0 \qquad \text{for } h \geqq a \tag{20}$$

and that

$$\eta(h) \begin{cases} <1 & (a \leqq h < b) \\ =1 & (h = b) \\ >1 & (h > b). \end{cases} \tag{21}$$

In (20), (21), and in what follows, h is understood to be restricted to the interior of the interval in which η exists.

Now choose and fix an $h \geqq a$ and define

$$dF_h = [\eta(h)]^{-1} e^{hx}\, dG. \tag{22}$$

Then the moment generating function of F_h, say ϕ, is given by

$$\phi(t) = \eta(t + h)/\eta(h).$$

Since $E(x \mid F_h) = \phi'(0) = \eta'(h)/\eta(h)$, it follows from (20) and the choice of h that $E(x \mid F_h) \geqq 0$. Consequently, by well-known properties of cumulative sums, $P(n < \infty \mid F_h) = 1$. Since $dG = \eta(h)e^{-hx}\, dF_h$ by (22), and

$$\eta(h) = [\phi(-h)]^{-1},$$

Theorem 1 yields the identity

$$P(n < \infty \mid G) = E(e^{-hs}[\eta(h)]^n \mid F_h), \tag{23}$$

valid for all $h \geqq a$.

Letting $h = b$ in (23), we have

$$P(n < \infty \mid G) = E(e^{-bs} \mid F_b), \tag{24}$$

by (21). Since $n < \infty$ implies $s \geqq c$, since $P(n < \infty \mid F_b) = 1$, and since $b > 0$, (24) yields

$$P(n < \infty \mid G) \leqq e^{-bc}, \tag{25}$$

which is de Finetti's inequality.

It is clear from the preceding derivation of (25) that the equality sign holds in (25) if and only if $P(s = c \mid F_b) = 1$ This condition can be shown to be equivalent to the condition that x be a discrete random variable taking only one positive value, say d, and that the negative values of x be integral multiples of d.[2] The condition is satisfied, in particular, if x takes only two values.

We turn now to the case when s can exceed c with positive probability. In this case, the effect of the 'excess over the boundary' can be estimated by means of an argument due to Wald [1][3] It is possible and advantageous to apply the argument to (23) rather than (24), as follows

Suppose that F_h obtains Write $y = c$ if $n = 1$, and

$$y = c - (x_1 + \cdots + x_{n-1})$$

if $1 < n < \infty$ Then y is well defined and $0 < y < \infty$ with probability 1. Let ξ denote the conditional expectation of e^{-hx_n} given n and y. It is not difficult to see that *ξ depends only on y and h; in fact*

$$\xi = E(e^{-hx} \mid x \geqq y, F_h). \tag{26}$$

We observe next that $s = c - y + x_n$ with probability one. Consequently, the right side of (23) can be written as $e^{-hc}\cdot E(e^{hy}\cdot\xi(y)\cdot\eta^n \mid F_h)$. It follows hence, by regarding y as a real variable confined to positive values, and setting

$$f(h) = \inf_y \{e^{hy}\cdot\xi\}, \qquad g(h) = \sup_y \{e^{hy}\cdot\xi\}, \tag{27}$$

that

$$e^{-hc} f(h)\, E(\eta^n \mid F_h) \leqq P(n < \infty \mid G) \leqq e^{-hc}\cdot g(h)\cdot E(\eta^n \mid F_h). \tag{28}$$

---

[2] In this case a slight extension of the methods of this paper can be used to obtain the probability distribution of n, with ∞ a possible value of n.

[3] Wald used the argument, in the context of a random walk with two absorbing barriers, to find the maximum possible effect of the excess over a barrier on the probability of absorption in that barrier However, the argument also yields the minimum possible effect, in Wald's context as well as in the present one.

Next, an easy calculation using (22), (26), and (27) shows that

$$f(h) = \inf_y \{e^{hy}/E(e^{hx} \mid x \geqq y, G)\},$$
(29)
$$g(h) = \sup_y \{e^{hy}/E(e^{hx} \mid x \geqq y, G)\}.$$

Finally, since $n \geqq 1$, we see from (21) and (28) that

(30) $$P(n < \infty \mid G) \leqq e^{-hc}\cdot g(h)\cdot \eta(h)$$

for $a \leqq h \leqq b$, and

(31) $$P(n < \infty \mid G) \geqq e^{-hc}\cdot f(h)\cdot \eta(h)$$

for $h \geqq b$.

The infimum of the right side of (30) with h restricted to $[a, b]$ gives, of course, the best upper bound obtainable by this method, while the supremum of the right side of (31) with h restricted to $[b, \infty]$ gives the best lower bound. In particular, taking $h = b$ in (30) and (31), we have

(32) $$e^{-bc}\cdot f(b) \leqq P(n < \infty \mid G) \leqq e^{-bc}\cdot g(b).$$

Another special bound is

(33) $$P(n < \infty \mid G) \leqq \inf_{a \leqq h \leqq b} \{e^{-hc}\cdot \eta(h)\};$$

this follows from (30) since $0 < g \leqq 1$.

It is easy to see from the preceding argument that in case x is bounded from above, η can be replaced by η^k in (30), (31) and (33), where k is the least positive integer such that $P(n = k \mid G) > 0$.

In concluding this section let us consider an example. In this example,

(34) $$G = pH + (1 - p)K$$

where $0 < p < 1$, H is some distribution function (possibly degenerate) confined to $(-\infty, 0]$, and

(35) $$dK(x) = \begin{cases} \lambda e^{-\lambda x}\, dx & \text{for} \quad x > 0 \\ 0 \text{ otherwise,} \end{cases}$$

where λ is a positive constant. It is assumed that H (and therefore G) admits a moment generating function in a neighbourhood of the origin, and that

$$E(x \mid G) \equiv pE(x \mid H) + (1 - p)/\lambda < 0.$$

It then follows that the equation

(36) $$\eta(b) \equiv pE(e^{bx} \mid H) + (1 - p)(\lambda/(\lambda - b)) = 1$$

has a unique non-zero solution b, with $0 < b < \lambda$.

A simple calculation, which is omitted, shows that in the present case we have $f(h) = g(h) = (\lambda - h)/\lambda$ for all h, where f and g are defined by (29). Consequently, it follows from (32) that

(37) $$P(n < \infty \mid G) = e^{-bc}\cdot(\lambda - b)/\lambda.$$

REFERENCES

[1] A. WALD, *Sequential Analysis*, John Wiley and Sons, Inc., New York (1947), pp. 157–164.

[2] A. WALD, "Differentiation under the expectation sign in the fundamental identity of sequential analysis", *Ann. Math. Stat.* Vol. 17 (1946), pp. 493–497.

[3] G. E. ALBERT, "A note on the fundamental identity of sequential analysis", *Ann. Math. Stat.*, Vol. 18 (1947), pp. 593–596.

[4] G. E. ALBERT, "Correction to 'A note on the fundamental identity of sequential analysis'," *Ann. Math. Stat* Vol. 19 (1948), p 426.

[5] D. BLACKWELL AND M. A. GIRSHICK, "On functions of sequences of independent chance vectors with applications to the problem of the "random walk" in k dimensions", *Ann Math Stat.*, vol 17 (1946), pp. 310–317.

[6] J. L. DOOB, *Stochastic processes*, John Wiley and Sons, Inc., New York (1953), pp 350–352.

[7] W. FELLER, *An introduction to probability theory and its applications*, John Wiley and Sons., Inc., New York (1950), Chap 14

[8] J. DUBOURDIEU, *Theorie mathématique du risque dans les assurances de repartition*, Gauthier-Villars, Paris (1952), pp 163–174.

[9] R. R. BAHADUR, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp 423–462, Section 8

[10] M. A. GIRSHICK AND L J. SAVAGE, "Bayes and minimax estimates for quadratic loss functions", *Proc. Second Berkeley Symp on Math. Stat and Probability*, University of California Press (1951), p 53

[11] B. O. KOOPMAN, "On distributions admitting a sufficient statistic," *Trans Amer Math. Soc.*, Vol 39 (1936), p 399

[12] A. WALD, "Some generalizations of the theory of cumulative sums of random variables," *Ann. Math Stat*, Vol. 16 (1945), pp 287–293

[13] M. A. GIRSHICK, "Contributions to sequential analysis II, III," *Ann Math Stat*, Vol. 17 (1946), pp 291–298.

[14] H. CHERNOFF, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann Math Stat*, Vol 23 (1952), pp. 493–507.

[15] P. HARTMAN AND A WINTNER, "On the law of the iterated logarithm", *Amer J. Math.*, Vol. 63 (1941), pp 169–176

[16] E. C. TITCHMARSH, *Theory of functions* Oxford Univ. Press, New York, (Second Edition, 1939).

ON THE STATISTICAL TREATMENT OF STOCHASTIC PROCESSES

By Isidore Fleischer and Anthony Kooharian

Bell Telephone Laboratories

1. Introduction. Grenander [3] has shown the feasibility of applying statistical techniques to stochastic processes. The basic tool involved in his approach is a representation theory for stochastic functions developed by Karhunen [4].

The present paper is an attempt to develop this approach on a systematic basis. Our analysis hinges on Theorem 2 and its corollary, a sharpened version of the Karhunen representation.

The statistical concepts we apply are the Neyman-Pearson criterion for hypothesis testing and the maximum likelihood estimate.

2. Basic theory. We shall require the following assumptions throughout the analysis:

(*) The stochastic function $x(t, \omega)$ is square integrable on Ω for each $t \varepsilon T$. The covariance function $r(s, t)$ is continuous in each variable and $r(t, t)$ is integrable on T.[1]

Although Karhunen [4] has established a representation theory for stochastic functions under much wider conditions, the following result (Karhunen [5]) suffices in our case:

Theorem 1. *If $x(t, \omega)$ satisfies (*) and has mean value $m(t) = 0$, then*

$$x(t, \omega) = \sum_{k=1}^{\infty} \frac{z_k(\omega)}{\sqrt{\lambda_k}} \varphi_k(t), \tag{1}$$

where the equality means convergence in the mean on Ω for each $t \varepsilon T$. The λ_k and φ_k are the eigenvalues and eigenfunctions, respectively, of the integral equation

$$\varphi_k(s) = \lambda_k \int_T r(s, t)\varphi_k(t)\, dt, \qquad k = 1, 2, \cdots, \tag{2}$$

and the $\{z_k\}$ are a set of mutually uncorrelated random variables with zero mean. In addition, the following relations hold:

$$E\{\overline{z_k(\omega)}x(t, \omega)\} = \varphi_k(t)/\sqrt{\lambda_k}, \qquad k = 1, 2, \cdots, \tag{3}$$

$$\int_T \overline{\varphi_k(t)}x(t, \omega)\, dt = z_k(\omega)/\sqrt{\lambda_k}, \qquad k = 1, 2, \cdots. \tag{4}$$

Received July 31, 1957; revised November 12, 1957.

[1] T may typically be taken as an arbitrary interval on the real line, although the entire analysis could be carried through considering T to be a topological space possessing a σ-finite Borel measure.

Remark 1. Equation (4) involves the definite integral of a stochastic function with respect to its "indexing" or t parameter. Unless stated to the contrary, such an integral is to be taken in the sense of Karhunen [4]; namely, that for arbitrary $u \in L_2(\Omega)$,

$$(4') \qquad \int_T \overline{\varphi_k(t)} E\{\bar{u}(\omega) x(t, \omega)\}\, dt = E\left\{\overline{u(\omega)}\, \frac{z_k(\omega)}{\sqrt{\lambda_k}}\right\}.$$

Remark 2. The representation (1) is actually the K-integral of the stochastic function $z(k)$ against φ_k over the indexing set $k = 1, 2, \cdots$ with measure

$$1/\sqrt{\lambda_k}.$$

In other terms the K-representation converges to $x(t)$ (Ω weakly) for each $t \in T$. We next prove

THEOREM 2. *The K-representation converges to $x(t)$ (Ω weakly) in the mean on T.*

Proof. By (*) the complex valued functions $E\{\bar{u}x(t)\}$ for arbitrary $u \in L_2(\Omega)$ are square integrable on T. In view of Remark 1 the Fourier coefficients of $E\{\bar{u}x(t)\}$ with respect to the orthonormal set $\{\varphi_k\}$ are precisely $E\{\bar{u}z_k/\sqrt{\lambda_k}\}$. The theorem then follows if it can be shown that for any square integrable $\varphi \perp \varphi_k$, $k = 1, 2, \cdots$,

$$I(u) \equiv \int_T E\{\bar{u}x(t)\}\overline{\varphi(t)}\, dt$$

vanishes identically for $u \in L_2(\Omega)$.

The linear functional I, however, is easily shown to be bounded which, together with the fact that by Mercer's theorem

$$I\{x(s)\} \equiv 0 \quad \text{for} \quad s \in T,$$

suffices.

COROLLARY. *If, in addition to (*), $x(t, \omega) \in L_2(T)$ for almost all $\omega \in \Omega$, then the $\{z_k\}$ defined by Eq. (4) may be identified with the ordinary Fourier coefficients of $x(t)$ with respect to $\{\varphi_k\}$, so that the K-representation converges in the mean on T wherever $x(t) \in L_2(T)$.*

Remark 3. The additional hypothesis involved in the corollary is satisfied if $x(t, \omega)$ is measurable on the product space $T \times \Omega$.

S. P. Lloyd has pointed out that this is the case if $x(t)$ is taken as one of its standard modifications ([2] p. 61–65).

3. Hypothesis testing for stochastic functions. In this section we shall apply the Neyman-Pearson test to stochastic functions satisfying the corollary.

Let the probability measure on Ω according to the null hypothesis be P_0 and according to the alternative be P_1. Then if $\{\lambda_k^0, \varphi_k^0\}$ and $\{\lambda_k^1, \varphi_k^1\}$ denote the eigenvalues and eigen-functions associated with the integral equations for the respective covariance functions, r_0 and r_1, we obtain the dual representations

$$(5) \qquad x(t, \omega) = \sum_{k=1}^{\infty} \frac{z_k^0(\omega)}{\sqrt{\lambda_k^0}} \varphi_k^0(t) + m^0(t)$$

and

(6) $$x(t, \omega) = \sum_{k=1}^{\infty} \frac{z_k^1(\omega)}{\sqrt{\lambda_k^1}} \varphi_k^1(t) + m^1(t)$$

in the sense of the corollary, where m^0 and m^1 are the respective means.

It may easily be shown that

(7) $$\frac{z_j^1}{\sqrt{\lambda_j^1}} = \sum_{k=1}^{\infty} a_{jk} \frac{z_k^0}{\sqrt{\lambda_k^0}} + b_j^1, \qquad j = 1, 2, \cdots,$$

almost everywhere P_0 and P_1, where

$$a_{jk} = \int_T \overline{\varphi_j^1}(t)\varphi_k^0(t)\, dt,$$

$$b_j^1 = \int_T \overline{\varphi_j^1}(t)[m^0(t) - m^1(t)]\, dt.$$

Assuming that the finite dimensional measures induced by P_0 and P_1 on the sample space of the z's are absolutely continuous with respect density functions g_N^0 and g_N^1, then the likelihood ratio is

(8) $$l_N = \frac{g_N^1(z_1^1, \cdots, z_N^1)}{g_N^0(z_1^0, \cdots, z_N^0)},$$

$$l_N = \frac{g_N^1\left(\sqrt{\lambda_1^1} \sum_{k=1}^{\infty} a_{1k} \frac{z_k^0}{\sqrt{\lambda_k^0}} + b_1^1, \cdots, \sqrt{\lambda_N^1} \sum_{k=1}^{\infty} a_{Nk} \frac{z_k^0}{\sqrt{\lambda_k^0}} + b_n^1\right)}{g_N^0(z_1^0, \cdots, z_N^0)}.$$

The martingale convergence theorem [2] then implies that l_N converges to the likelihood ratio on the infinite sample space almost everywhere P_0. Hence

THEOREM 3. *The Neyman-Pearson critical region for the rejection of P_0 against P_1 is*

$$\lim_{N\to\infty} l_N \geqq k.$$

An application. It follows from Eq. (4) that if $x(t)$ is a normal process, then the z_i are normally distributed as well. Since the z_i are uncorrelated they are now independent so that their finite dimensional densities are explicitly known in terms of the eigenvalues.

Thus in testing for the covariance function in the normal case with zero mean, the appropriate likelihood ratio is the limit of

(9) $$l_N = \sqrt{\frac{\lambda_1^1 \cdots \lambda_N^1}{\lambda_1^0 \cdots \lambda_N^0}} \exp\left\{\frac{1}{2}\left[\sum_{k=1}^{N} \lambda_k^0 (z_k^0)^2 - \lambda_k^1 \left(\sum_{j=1}^{\infty} \frac{a_{jk} z_j^0}{\sqrt{\lambda_j^0}}\right)^2\right]\right\}.$$

As an illustration consider the case where it is desired to test the hypothesis that a stationary normal process on the interval $(0, 2\pi)$ with mean zero has the covariance function.

$$r_0(\tau) = \sum_{j=-\infty}^{\infty} \alpha_j e^{ij\tau}, \qquad i = \sqrt{-1},$$

against the alternative

$$r_1(\tau) = \sum_{j=-\infty}^{\infty} \beta_j e^{ij\tau}$$

Here we must have

$$0 \leq \alpha_j, \beta_j, \qquad \sum_{j=-\infty}^{\infty} \alpha_j = \sum_{j=-\infty}^{\infty} \beta_j = 1$$

The reader will readily convince himself that

$$\varphi_k^0(t) = \varphi_k^1(t) = e^{ikt}, \qquad \lambda_k^0 = \frac{1}{\alpha_k}, \qquad \lambda_k^1 = \frac{1}{\beta_k}$$

It follows that the matrix transforming the z^0 into the z^1 is the identity matrix $a_{jk} = \delta_{jk}$. The likelihood ratio is thus the limit of

$$(10) \qquad l_N(z_1^0, \cdots, z_N^0) = \sqrt{\frac{\alpha_1 \cdots \alpha_N}{\beta_1 \cdots \beta_N}} \exp\left\{\frac{1}{2}\sum_{k=1}^{N} (z_k^0)^2 \left(\frac{\beta_k^2 - \alpha_k^2}{\alpha_k \beta_k^2}\right)\right\}$$

Since the logarithm is monotone and continuous, we may describe the critical region by

$$\lim_{N\to\infty} \log \prod_{i=1}^{N} \gamma_i + \sum_{k=1}^{N} \frac{(z_k^0)^2}{\alpha_k}(1 - \gamma_k^2) \geq K,$$

where $\gamma_k = \alpha_k/\beta_k$. If $\prod_{i=1}^{\infty} \gamma_i$ exists this may also be written

$$(11) \qquad \sum_{k=1}^{\infty} \lambda_k^0 (z_k^0)^2 (1 - \gamma_k^2) \geq K^*$$

4. Estimation theory for stochastic functions. In this section we shall treat the problem of applying statistical estimation to stochastic functions. We shall show how the classical maximum likelihood estimate can be adapted to the situation at hand, and illustrate by an example.

Let $x(t)$ be a stochastic function with underlying probability space Ω and suppose that the measure on Ω is known to be one of a family, P_α, $\alpha \in A$. It is desired to estimate α on the basis of a single observation of $x(t)$.

We shall assume that each P_α is absolutely continuous with respect to some fixed probability measure μ. Thus there exists a family of positive μ-measurable functions $f(\omega, \alpha)$ such that for every measurable set S in Ω,

$$(12) \qquad P_\alpha(S) = \int_S f(\omega, \alpha)\, d\mu.$$

In these circumstances it is appropriate to require that in addition to (*), $x(t)$ be square integrable in t almost everywhere μ. Since we shall commit our-

selves to using only the L^2 properties of our realization, we may as well assume that $f(\omega, \alpha)$ has the same value on realizations equal almost everywhere T.[2]

Each P_α generates a Karhunen representation

$$(13) \qquad x(t) = \sum_{i=1}^{\infty} \frac{z_i(\omega, \alpha)}{\sqrt{\lambda_i(\alpha)}} \varphi_i(t, \alpha) + m_\alpha(t).$$

Here the z_i have a distribution known in terms of P_α. Furthermore, for different α, these variables are related analogously to (7). We have, in fact,

$$(14) \qquad \frac{z_j(\omega, \beta)}{\sqrt{\lambda_j(\beta)}} = \sum_{i=1}^{\infty} \frac{z_i(\omega, \alpha)}{\sqrt{\lambda_i(\alpha)}} a_{ij}(\alpha, \beta) + \int_T [m^\alpha(t) - m^\beta(t)] \overline{\varphi_j(t, \beta)}\, dt,$$

where

$$a_{ij}(\alpha, \beta) = \int_T \varphi_i(t, \alpha) \overline{\varphi_j(t, \beta)}\, dt$$

and the equality holds almost everywhere μ.

Let $f_N(z, \cdots, z_N; \alpha)$ be the density function of P_α with respect to μ on the sample space of $z_i(\omega, \alpha)$. Again by the martingale convergence theorem [2], there follows:

THEOREM 4. *If A is separable and if $f(\omega, \alpha)$ depends continuously on α almost everywhere μ, then $f(\omega, \alpha)$ and, therefore, the maximum likelihood estimate of α can be calculated from the f_N.*

Example. Let $x(t)$ be a normally distributed stochastic function with known covariance whose mean is to be estimated. We arbitrarily choose the dominating measure μ to be that one corresponding to $m(t) = 0$.

Since

$$x_j = \frac{z_j}{\sqrt{\lambda_j}} + m_j^\alpha,$$

the finite dimensional density function for the x_j may be expressed in the form

$$g_N(x_1, \cdots, x_N; \alpha) = \frac{\sqrt{\lambda_1 \cdots \lambda_N}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2} \sum_{j=1}^{n} \lambda_j (x_j - m_j^\alpha)^2\right\}$$

so that the likelihood ratio is

$$(15) \qquad \begin{aligned} \lim_{N\to\infty} f_N(x_1, \cdots, x_N; \alpha) &= \lim_{N\to\infty} \exp\left\{\frac{1}{2} \sum_{j=1}^{N} \lambda_j [x_j^2 - (x_j - m_j^\alpha)^2]\right\} \\ &= \exp\left\{\frac{1}{2} \sum_{j=1}^{\infty} \lambda_j [2 m_j^\alpha x_j - (m_j^\alpha)^2]\right\}. \end{aligned}$$

The likelihood ratio clearly achieves its maximum for

$$m_j^a = x_j, \qquad j = 1, 2, \cdots.$$

[2] What this amounts to, in effect, is to replace f by $E\{f \mid z_1, z_2, \cdots\}$.

The only possible maximum likelihood estimate is, therefore,

$$m^*(t) = x(t).$$

It may well happen, however, that the realization $x(t)$ cannot serve as a mean. For example, it follows from (*) that $m^*(t)$ is continuous for every α so that if the observed realization does not have this property, the maximum likelihood estimate does not exist.

BIBLIOGRAPHY

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton, 1951
[2] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, New York, 1953
[3] U. Grenander, "Stochastic processes and statistical inference," *Arkiv för Mat.*, Vol. 1 (1950), pp. 195–277.
[4] K. Karhunen, "Über lineare Methoden in der Wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae*, Series A, I 37, Helsinki (1947), pp. 1–79.
[5] K. Karhunen, "Zur Spektraltheorie stochasticher Prozesse", *Ann. Acad. Sci. Fennicae*, Series A, I 34, Helsinki (1946), pp. 1–7.

SOME REMARKS ON SAMPLING WITH REPLACEMENT

By Des Raj and Salem H. Khamis

American University of Beirut

1. Introduction and summary. In order to estimate the mean of a finite population from a random sample, the sample units may be selected in two ways. A pre-determined number m of units may be selected with replacement or sampling with replacement may be continued till a desired number n of distinct units is obtained. The first procedure is called sampling with replacement while the second may be called sampling without replacement. A comparison is generally made between the two procedures with $m = n$. This comparison, however, is not fair since costs are usually proportional to the number of distinct units and in the first procedure this number would be less than or equal to n.

In the first procedure the population mean is usually estimated by the sample mean based on all the units in the sample including repetitions while in the second procedure the estimate is generally made to depend on the distinct units only. The object of this paper is to show that the estimate making use of only the distinct units is superior in either procedure. The following results are proved in this paper:

(i) In sampling with replacement the estimate of the mean based on distinct units in the sample is superior to the estimate based on the total sample size when (a) the total sample size is fixed in advance, while the number of distinct units in the sample is a random variable, and, (b) when the total sample size is a random variable while the number of distinct units is fixed in advance.

(ii) The same is true of ratio estimates. It is also shown that the bias is numerically less if the ratio estimate is based on distinct units regardless of whether these are fixed in advance or considered as random variables.

(iii) Expressions for the estimation of the variances of the various estimates considered in this paper are given.

(iv) The above results are extended to multistage sampling.

2. Statement of the problem. Let us consider a finite population consisting of N sampling units. Suppose we are interested in estimating the population mean $\bar{Y}$ for a character y, from a sample selected with replacement with equal probabilities. We consider the following two sampling schemes.

Scheme A. We select with replacement a total sample of size m fixed in advance. We denote the number of distinct sample units selected by the random variable u.

Scheme B. We select with replacement a sample of n distinct units fixed in advance. We denote the total sample size, including repetitions, by the random variable v.

Received February 15, 1957; revised November 18, 1957.

We consider Scheme A first. Let $a_1, a_2, \cdots, a_u$ be the u distinct units selected in the sample and let k_r be the number of times the rth unit a_r occurs in the sample, with $\sum k_r = m$. We compare the bias and variances of the usual estimate

$$\bar{y}_m = \frac{1}{m} \sum_{r=1}^{u} k_r y_r , \tag{1}$$

and the estimate

$$\bar{y}_u = \frac{1}{u} \sum_{r=1}^{u} y_r \tag{2}$$

based on the distinct units only, where y_r is the value of the character y for the rth distinct unit in the sample.

The estimate (1) is well known to be unbiased and that its variance is given by

$$V(\bar{y}_m) = \frac{1}{m}\left(1 - \frac{1}{N}\right)\sigma^2 = \left(Q - \frac{1}{N}\right)\sigma^2 \tag{3}$$

where

$$Q = \frac{1}{m}\left(1 + \frac{m-1}{N}\right) \tag{4}$$

and

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2. \tag{5}$$

For a given u the expected value of $\bar{y}_u$ is $\bar{Y}$ so that $\bar{y}_u$ is an unbiased estimate of $\bar{Y}$. With regard to the variance of $\bar{y}_u$ we have

$$V(\bar{y}_u) = \left[E\left(\frac{1}{u}\right) - \frac{1}{N}\right]\sigma^2. \tag{6}$$

The estimate $\bar{y}_u$ is superior to $\bar{y}_m$ if

$$E\left(\frac{1}{u}\right) < Q. \tag{7}$$

The probability distribution of the random variable u is given by (cf. Feller, [1])

$$P(u) = N^{-m} \binom{N}{u} \Delta^u 0^m, \tag{8}$$

where the sth difference of 0^t is defined by

$$\Delta^s 0^t = \sum_{r=0}^{s} (-1)^{s-r} \binom{s}{r} r^t.$$

Hence

$$E\left(\frac{1}{u}\right) = N^{-m} \sum_{u=1}^{m} \frac{1}{u} \binom{N}{u} \Delta^u 0^m \tag{9}$$

and the expected sample size is

$$(10) \qquad E(\nu) = N^{-r} \sum_{u=1}^{r} u \binom{N}{u} \Delta^u 0^r.$$

We consider now Scheme B. Let $b_1, b_2, \cdots, b_n$ be the n distinct units selected in the sample and let k_r be the number of times the rth unit b_r occurs in the sample with $\sum k_r = r$. We compare the bias and variances of the estimate

$$(11) \qquad \bar{y}_r = \frac{1}{r} \sum_{r=1}^{n} k_r y_r$$

and the estimate

$$(12) \qquad \bar{y}_n = \frac{1}{n} \sum_{r=1}^{n} y_r$$

where y_r is the same as in Scheme A with obvious modifications. It is easy to see that the estimates $\bar{y}_r$ and $\bar{y}_n$ are unbiased for estimating $\bar{Y}$. Also we have

$$(13) \qquad V(\bar{y}_r) = E\left(\frac{1}{\nu}\right)\frac{N-1}{N} \sigma^2$$

and

$$(14) \qquad V(\bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right) \sigma^2.$$

The estimate $\bar{y}_n$ is superior if

$$(15) \qquad E\left(\frac{1}{\nu}\right) > \frac{1}{n} \frac{N-n}{N-1}.$$

The probability distribution of the random variable ν can be shown to be

$$(16) \qquad P(\nu) = \binom{N-1}{n-1} N^{1-\nu} \Delta^{n-1} 0^{\nu-1}.$$

Hence

$$(17) \qquad E\left(\frac{1}{\nu}\right) = \binom{N-1}{n-1} \sum_{\nu=n}^{\infty} \frac{1}{\nu} N^{1-\nu} \Delta^{n-1} 0^{\nu-1}.$$

We give in Table 1 a numerical table for selected sample sizes which illustrates the numerical magnitudes of the differences discussed above. The theoretical proofs are given in Section 3 below.

3. Proofs of the inequalities. In order to establish inequality (7), we shall first prove the following

LEMMA. *Let*

$$(18) \qquad S_{t,N} = \sum_{u=1}^{m-t} \frac{1}{u+t} \binom{N}{u} \Delta^u 0^{m-t}.$$

TABLE 1

Values of $E(u)$, $E\left(\frac{1}{u}\right)$, Q and $1/E(u)$

N	m	$E(u)$	$E\left(\frac{1}{u}\right)$	Q	$\frac{1}{m}$	$\frac{1}{E(u)}$
$N = 10$	1	1.00	1 000	1.000	1 000	1.000
	2	1.90	550	550	500	526
	3	2.71	.385	.400	333	369
	4	3.44	303	325	250	291
	5	4 10	253	280	200	244
	6	4.69	.221	.250	167	213
$N = 50$	1	1 00	1 000	1 000	1 000	1 000
	2	1.98	.510	.510	.500	505
	3	2.94	.343	347	333	340
	4	3 88	260	.265	.250	258
	5	4.80	210	.216	200	.208
	6	5.71	.177	.183	167	.175
$N = 100$	1	1.00	1.000	1.000	1.000	1.000
	2	1.99	.505	.505	.500	.503
	3	2.97	338	.310	.333	.337
	4	3.94	.255	.258	.250	254
	5	4.90	.205	.208	.200	.204
	6	5.85	.172	.175	.167	.171

Then

(19) $\qquad S_{t,N} \leq (N+1)S_{t+1,N}$ *for* $t = 0, 1, \cdots, m-1; \qquad m > 2,$

the sign of equality holds only for $t = 0$.

Proof.

$$S_{t,N} = \sum_{u=1}^{m-t} \frac{u}{u+t}\binom{N}{u}(\Delta^{u-1}0^{m-t-1} + \Delta^u 0^{m-t-1})$$

$$= \sum_{u=2}^{m-t} \frac{N-u+1}{u+t}\binom{N}{u-1}\Delta^{u-1}0^{m-t-1} + \sum_{u=1}^{m-t-1} \frac{u}{u+t}\binom{N}{u}\Delta^u 0^{m-t-1}$$

$$= \sum_{u=1}^{m-t-1}\left(\frac{N-u}{u+t+1} + \frac{u}{u+t}\right)\binom{N}{u}\Delta^u 0^{m-t-1}$$

$$= \sum_{u=1}^{m-t-1} \frac{(N+1)u + tN}{(u+t)(u+t+1)}\binom{N}{u}\Delta^u 0^{m-t-1}.$$

Thus $S_{0,N} = (N+1)S_{1,N}$, and $S_{t,N} < (N+1)S_{t+1,N}$ for $1 \leq t \leq m-1$. This proves the lemma.

Corollary. Applying the lemma $m-2$ times beginning with $t = 1$, we have, for $m > 2$

$$S_{1,N} < (N+1)^{m-2}S_{m-1,N} = (N+1)^{m-2}\frac{N}{m}$$

so that

(20) $$S_{1,N-1} < N^{m-2}\frac{N-1}{m}.$$

Now, inequality (7) is proved in the following

THEOREM.

(21) $$E\left(\frac{1}{u}\right) \leqq \frac{1}{N} + \frac{N-1}{N}\frac{1}{m};$$

the sign of equality holds only for $m = 2$.

PROOF.

$$\begin{aligned} E\left(\frac{1}{u}\right) &= N^{-m}\sum_{u=1}^{m}\binom{N}{u}(\Delta^u 0^{m-1} + \Delta^{u-1}0^{m-1}) \\ &= \frac{1}{N} + N^{1-m}\sum_{2}^{m}\frac{1}{u}\binom{N-1}{u-1}\Delta^{u-1}0^{m-1} \\ &= \frac{1}{N} + N^{1-m}\sum_{1}^{m-1}\frac{1}{u+1}\binom{N-1}{u}\Delta^u 0^{m-1} \\ &= \frac{1}{N} + N^{1-m}S_{1,N-1}. \end{aligned}$$

On using (20) we have for $m > 2$,

$$E\left(\frac{1}{u}\right) < \frac{1}{N} + \frac{N-1}{N}\frac{1}{m}.$$

For $m = 2$, $S_{1,N-1} = \frac{1}{2}(N-1)$, so that the last inequality reduces to an equality in this case.

To prove inequality (15), we make use of the following

LEMMA. *If v is a positive random variable, then*

(22) $$E\left(\frac{1}{v}\right) \geqq \frac{1}{E(v)}.$$

The proof of this lemma follows from Cauchy's inequality (cf. Hardy and others, [3]),

(23) $$\left(\sum a^2\right)\left(\sum b^2\right) > \left(\sum ab\right)^2,$$

by substituting $a = \sqrt{vP(v)}$, $b = \sqrt{v^{-1}P(v)}$, and noting that the two sides of the inequality are convergent because $E(v)$ and $E(1/v)$ are finite.

Now (cf. Feller, [2])

$$E(v) = N\left(\frac{1}{N} + \frac{1}{N-1} + \cdots + \frac{1}{N-n+1}\right),$$

$$\therefore E\left(\frac{1}{v}\right) > \frac{1}{N\dfrac{n}{N-n+1}} = \frac{N-n+1}{nN} > \frac{N-n}{n(N-1)} \quad \text{for} \quad n > 1.$$

It is easy to see that for $n = 1$ the two procedures are equivalent and the inequality (15) reduces to an equality.

4. Estimation of variances. We shall consider the problem of estimating from the sample the variances of $\bar{y}_u$ and $\bar{y}_v$. Under Scheme A, it is easy to see that for a given $u \geqq 2$, an unbiased estimate of σ^2 is provided by

$$s_u^2 = \frac{1}{u-1} \sum_{i=1}^{u} (y_i - \bar{y}_u)^2. \tag{24}$$

Thus considering

$$G_u = \left[\left(\frac{1}{u} - \frac{1}{N}\right) + N^{1-n}\left(1 - \frac{1}{u}\right)\right] s_u^2, \tag{25}$$

we have

$$E[G_u \mid u \geqq 2] = V(\bar{y}_u); \tag{26}$$

so that G_u provides an unbiased estimate of the variance of $\bar{y}_u$. It is unbiased in the conditional sense, namely when the number of distinct units in the sample exceeds unity. An alternative unbiased estimate is provided by G_u' where

$$G_u' = \left[\left(\frac{1}{u} - \frac{1}{n}\right) + \frac{N-1}{N^n - N}\right] s^2,$$

(27)

$$s^2 = \frac{1}{u-1} \sum_{i=1}^{u} (y_i - \bar{y}_u)^2, \qquad \text{for } u \geqq 2,$$

and

$$s^2 = 0, \qquad \text{for } u = 1.$$

Under Scheme B, it is easy to see that

$$E\left[\frac{1}{v} s_v^2 \mid v\right] = \frac{1}{v} \frac{N-1}{N} \sigma^2 \tag{28}$$

where

$$s_v^2 = \frac{1}{v-1} \sum_{i=1}^{v} (y_i - \bar{y}_v)^2 \tag{29}$$

Thus

$$E\left[\frac{1}{v} s_v^2\right] = E\left(\frac{1}{v}\right) \frac{N-1}{N} \sigma^2, \tag{30}$$

so that $1/v\, s_v^2$ is an unbiased estimate of $V(\bar{y}_v)$.

5. Extension to ratio estimation. We shall now extend the above results to ratio estimates. We make use of the notation and approximate results given by Cochran [1]. The object is to estimate the population ratio $R = Y/X$. Under Scheme A we compare the two estimates

$$R_m = \sum_{i=1}^{n} k_i y_i \Big/ \sum_{i=1}^{n} k_i x_i \tag{31}$$

and

(32) $$\bar{R}_u = \sum_{i=1}^{u} y_i \Big/ \sum_{i=1}^{u} x_i.$$

We have

(33)
$$V(\bar{R}_m) \doteq R^2\left(Q - \frac{1}{N}\right)[C_{yy} + C_{xx} - 2C_{xy}]$$
$$B(\bar{R}_m) \doteq R\left(Q - \frac{1}{N}\right)[C_{xx} - C_{xy}]$$
$$V(\bar{R}_u) \doteq R^2\left[E\left(\frac{1}{u}\right) - \frac{1}{N}\right][C_{yy} + C_{xx} - 2C_{xy}]$$
$$B(\bar{R}_u) \doteq R\left[E\left(\frac{1}{u}\right) - \frac{1}{N}\right][C_{xx} - C_{xy}]$$

where $B(\bar{R})$ stands for the absolute value of the bias of the estimate $\bar{R}$. Using inequality (7), we have

(34) $$V(\bar{R}_u) < V(\bar{R}_m) \text{ and } B(\bar{R}_u) < B(\bar{R}_m).$$

Under Scheme B, the estimates to be compared are

(35) $$\bar{R}_v = \sum_{i=1}^{n} k_i y_i \Big/ \sum_{i=1}^{n} k_i x_i$$

and

(36) $$\bar{R}_n = \sum_{i=1}^{n} y_i \Big/ \sum_{i=1}^{n} x_i.$$

It is easy to see that the estimate $\bar{R}_n$ is superior to $\bar{R}_v$ from the point of view of variance and bias.

6. Extension to multistage designs. The result obtained for unistage designs will now be extended to multistage designs. Let a population consist of N first stage sampling units, of which m or n are selected with equal probabilities with replacement according to the Schemes A or B respectively. For the ith first stage unit, let t_i (based on sampling at second and subsequent stages) be an unbiased estimate of y_i, the total value of the character y for the unit. For Schemes A and B the unbiased estimates considered are

(37) $$\bar{y}'_u = \frac{1}{u}\sum_{i=1}^{n} t_i,$$

(38) $$\bar{y}'_m = \frac{1}{m}\sum_{i=1}^{u} k_i t_i,$$

and

(39) $$\bar{y}'_n = \frac{1}{n}\sum_{i=1}^{n} t_i,$$

(40) $$\bar{y}'_v = \frac{1}{v}\sum_{i=1}^{n} k_i t_i.$$

The variances of the estimates are given by

(41) $$V(\bar{y}'_u) = \left[\left(1 + \frac{\delta^2}{\sigma^2}\right) E\left(\frac{1}{u}\right) - \frac{1}{N}\right]\sigma^2$$

(42) $$V(\bar{y}'_m) = \left[\frac{\delta^2}{\sigma^2} Q + \frac{N-1}{mN}\right]\sigma^2$$

(43) $$V(\bar{y}'_n) = \frac{\delta^2}{n} + \left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2,$$

(44) $$V(\bar{y}'_v) = \frac{N-1}{N} E\left(\frac{1}{v}\right)(\delta^2 + \sigma^2) + \frac{\delta^2}{N},$$

where

$$\delta^2 = \frac{1}{N}\sum_{1}^{N} V(t_i).$$

Using inequalities (7) and (15) it is found that $\bar{y}'_u$ is superior to $\bar{y}'_m$ while $\bar{y}'_n$ is superior to $\bar{y}'_v$.

REFERENCES

[1] W G. Cochran, "*Sampling Techniques*", John Wiley and Sons, New York, 1953, pp. 117, 118.

[2] W. Feller, "*An introduction to probability theory and its applications*", John Wiley and Sons, New York, 1950, pp. 69, 175.

[3] G. H. Hardy, J. E. Littlewood, and G. Polya, "*Inequalities*", Cambridge University Press, 1934, pp 16, 124.

It is easily verified that $\hat{p}$ is an unbiased estimate of p and

$$p = \int_{-\infty}^{+\infty} G(s)\, dF(s),$$

$$\hat{p} = \int_{-\infty}^{+\infty} G_n(s)\, dF_m(s),$$

hence

$$(2.4)\qquad \begin{aligned} p - \hat{p} &= \int_{-\infty}^{+\infty} G\, d(F - F_m) + \int_{-\infty}^{+\infty} (G - G_n)\, dF_m \\ &= \int_{-\infty}^{+\infty} (F_m - F)\, dG + \int_{-\infty}^{+\infty} (G - G_n)\, dF_m \end{aligned}$$

and

$$(2.5)\qquad p - \hat{p} \leqq D_m^- + D_n^+,$$

where

$$D_m^- = \sup_{-\infty < s < +\infty} \{F_m(s) - F(s)\},$$

$$D_n^+ = \sup_{-\infty < s < +\infty} \{G(s) - G_n(s)\}.$$

It is well known [2] that

$$\Pr\{D_m^- < v\} = \Pr\{D_m^+ < v\} = P_m(v)$$

and

$$\Pr\{D_n^+ < v\} = P_n(v)$$

are cumulative distribution functions which depend on the sample sizes m, n, but not on the c.d.f.'s F and G. It follows from (2.5) that

$$(2.6)\qquad \Pr\{p \leqq \hat{p} + \epsilon\} \geqq \Pr\{D_m^+ + D_n^+ \leqq \epsilon\} = P_{m,n}(\epsilon),$$

where $P_{m,n}(\epsilon)$ is the convolution of P_m and P_n, hence does not depend on F and G. The statistic $\hat{p}$ has, therefore, the property required of ψ in (2.2) provided one can, for given ϵ, α, determine numbers $M_{\epsilon,\alpha}$, $N_{\epsilon,\alpha}$ so that

$$(2.7)\qquad P_{m,n}(\epsilon) \geqq 1 - \alpha \qquad \text{for } m \geqq M_{\epsilon,\alpha},\ n \geqq N_{\epsilon,\alpha}.$$

Some further properties of $\hat{p}$ are discussed in [1].

A numerical procedure for computing $M_{\epsilon,\alpha}$, $N_{\epsilon,\alpha}$ is presented in the next sections.

3. An approximate expression for $P_{m,n}(\epsilon)$. It was shown by N. Smirnov [3] that

$$(3.1)\qquad \lim_{n\to\infty} \Pr\{D_n^+ \leqq z/\sqrt{n}\} = \lim_{n\to\infty} P_n(z/\sqrt{n}) = 1 - e^{-2z^2} = L(z).$$

Since, for fixed n, $P_n(z/\sqrt{n}) = H_n(z)$ is a cumulative distribution function, and $L(z) = 1 - e^{-2z^2}$ is a continuous c.d.f., it follows by a well-known argument (see, e.g., [4], p. 276) that $H_n(z) \to L(z)$ uniformly. We may, therefore, conclude that

$$\lim_{n\to\infty} [\Pr\{D_n^+ \leqq v\} - L(v\sqrt{n})] = \lim_{n\to\infty} [H_n(v\sqrt{n}) - L(v\sqrt{n})] = 0 \tag{3.2}$$

uniformly for $0 \leqq v \leqq 1$. Writing

$$\begin{aligned} P_{m,n}(\epsilon) &= \Pr\{D_m^+ + D_n^+ \leqq \epsilon\} = \int_0^\epsilon P_n(\epsilon - u)\, dP_m(u), \\ Q_{m,n}(\epsilon) &= \int_0^\epsilon L[(\epsilon - u)\sqrt{n}]\, dL(u\sqrt{m}), \end{aligned} \tag{3.3}$$

we have

$$\begin{aligned} | P_{m,n}(\epsilon) - Q_{m,n}(\epsilon) | &\leqq \left| \int_0^\epsilon \{P_n(\epsilon - u) - L[(\epsilon - u)\sqrt{n}]\}\, dP_m(u) \right| \\ &\quad + \left| \int_0^\epsilon \{P_m(\epsilon - v) - L[(\epsilon - v)\sqrt{m}]\}\, dL(v\sqrt{n}) \right| \\ &\leqq \operatorname*{Max}_{0\leqq u\leqq\epsilon} | P_n(\epsilon - u) - L[(\epsilon - u)\sqrt{n}] | \\ &\quad + \operatorname*{Max}_{0\leqq v\leqq\epsilon} | P_m(\epsilon - v) - L[(\epsilon - v)\sqrt{m}] |, \end{aligned} \tag{3.4}$$

which in view of (3.2) shows that

$$\lim_{\substack{m\to\infty \\ n\to\infty}} | P_{m,n}(\epsilon) - Q_{m,n}(\epsilon) | = 0 \tag{3.5}$$

uniformly for $0 \leqq \epsilon \leqq 1$. This justifies the use of $Q_{m,n}(\epsilon)$ as an approximation to $P_{m,n}(\epsilon)$ for m, n sufficiently large. Some observations on the goodness of this approximation are presented in Section 5.

By straightforward integration one obtains for $Q_{m,n}(\epsilon)$ the expression

$$\begin{aligned} Q_{m,n}(\epsilon) &= 1 - \frac{n}{m+n} e^{-2m\epsilon^2} - \frac{m}{m+n} e^{-2n\epsilon^2} \\ &\quad - \frac{2\sqrt{2\pi}\, mn\epsilon^2}{(m+n)^{3/2}} e^{-2mn\epsilon^2/(m+n)} \frac{1}{\sqrt{2\pi}} \int_{-2n\epsilon/\sqrt{m+n}}^{2m\epsilon/\sqrt{m+n}} e^{-t^2/2}\, dt. \end{aligned} \tag{3.6}$$

4. Sample sizes m, n which satisfy $Q_{m,n}(\epsilon) = 1 - \alpha$. With the notations

$$\begin{aligned} m + n &= N \\ m / (m+n) = \lambda, &\qquad n / (m+n) = 1 - \lambda \\ \epsilon\sqrt{m+n} &= \delta \end{aligned} \tag{4.1}$$

TABLE I

Values $\delta_{\lambda,\alpha}$ such that $Q(\delta_{\lambda,\alpha}\,;\lambda) = 1 - \alpha$

α	λ				
	.1	.2	.3	.4	.5
.10	4.1185	3.2027	2 8501	2.6928	2.6468
.05	4.6115	3 5667	3.1641	2 9844	2.9317
.01	5.5700	4.2745	3 7770	3.5524	3.4870
.005	5.9300	4.5405	4 0050	3 7665	3.6960
.001	6.6800	5 0980	4.4880	4 2150	4.1360

the expression (3.6) for $Q_{m,n}(\epsilon)$ may be written in the form

$$Q(\delta;\lambda) = 1 - \lambda e^{-2(1-\lambda)\delta^2} - (1-\lambda)e^{-2\lambda\delta^2} - 2\sqrt{2\pi}\,\lambda(1-\lambda)\,\delta e^{-2\lambda(1-\lambda)\delta^2}\frac{1}{\sqrt{2\pi}}\int_{-2(1-\lambda)\delta}^{2\lambda\delta} e^{-t^2/2}\,dt. \tag{4.2}$$

Table I contains solutions $\delta_{n,\alpha}$ of the equation.

$$Q(\delta;\lambda) = 1 - \alpha \tag{4.3}$$

for $\alpha = .001, .005, .01, .05, .10$, and $\lambda = .1\ (.1)\ .5$. These solutions were obtained on a desk calculator, using the National Bureau Bureau of Standards Tables of the Exponential Function [5], Descending Exponential [6], and the Normal Distribution Function [7].

The use of the quantities N, λ, δ instead of the original m, n, ϵ has not only the advantage of reducing the computations to a table with double entry, but also makes it possible to design an experiment with a given ratio $\lambda = m/N$. This ratio is often dictated by considerations of cost or time.

Example. We wish to use four times as many Y's as X's, i.e., $\lambda = .2$, and require $\epsilon = .10$, $\alpha = .05$. From Table I we have $\delta_{.2,.05} = 3.5667$; hence, by (4.1), $(.10)\sqrt{N} = 3\ 5667$, and $N = 1272.13$, $m = 254.43$, $n = 1017.70$. The rounded-up sample sizes are therefore 255 for X, 1018 for Y.

5. Concluding remarks. The sample sizes computed for given λ, ϵ, α by the use of Table I are conservative, i.e., too large, for two reasons. The first is that, instead of finding sample sizes m, n such that $P\{p \leqq \hat{p} + \epsilon\} = 1 - \alpha$, we used inequality (2.6) and looked for m, n satisfying $P_{m,n}(\epsilon) = 1 - \alpha$, a step which certainly yields larger values. The second reason is that in equation $P_{m,n}(\epsilon) = 1 - \alpha$ the exact expression $P_{m,n}(\epsilon)$ was replaced by the approximate expression $Q_{m,n}(\epsilon)$ and then only m, n were computed. This step was justified by (3.5) which, however, does not indicate which way the sample sizes are affected. The following arguments are offered in favor of the contention that the solutions m, n of $P_{m,n}(\epsilon) = 1 - \alpha$ differ little from those of $Q_{m,n}(\epsilon) = 1 - \epsilon$ and that the solu-

tions of the second equation are more conservative (greater) than those of first.

The exact form of $P\{D_n^+ \leqq v\}$ for finite n is known and numerical compu tions, some of which are reproduced in [8], show that already for $n \geqq 50$ approximation of $P\{D_n^+ \leqq v\}$ by $L(v\sqrt{n})$ is uniformly very good. Since the sa ple sizes computed from Table I are in all practical situations much larger t 50, (3.4) assures very close agreement between $P_{m,n}(\epsilon)$ and $Q_{m,n}(\epsilon)$.

Furthermore, the following conjecture appears to be substantiated by c siderable numerical computations and some analytical considerations, altho no proof for it is available: for every integer $n \geqq 1$ and for $0 \leqq v \leqq 1$,

$$L(v\sqrt{n}) = 1 - e^{-2nv^2} \leqq P\{D_n^+ \leqq v\}. \tag{5.1}$$

From (5.1) would follow that $P_{m,n}(\epsilon) \geqq Q_{m,n}(\epsilon)$ for $0 \leqq \epsilon \leqq 1$, hence $Q_{m,n}(\epsilon$ $1 - \alpha$ would yield sample sizes larger than $P_{m,n}(\epsilon) = 1 - \alpha$.

REFERENCES

[1] Z. W. BIRNBAUM, "On a use of the Mann-Whitney statistic," *Proceedings of the T Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, Univer of California Press, 1956, pp. 13–17.

[2] A. WALD AND J. WOLFOWITZ, "Confidence limits for continuous distribution functio *Ann. Math. Stat.*, Vol. 10 (1939), pp. 105–118.

[3] N. SMIRNOV, "Sur les écarts de la courbe de distribution empirique," *Recueil M (Mat. Sbornik)*, N.S., Vol. 6 (1939), pp. 3–26.

[4] M. FRÉCHET, "Generalités sur les probabilités. Variables aléatoires," *Traité de ca des probabilités* (edited by E. Borel), Gauthier-Villars, Paris, 1937.

[5] National Bureau of Standards, "Tables of the exponential function," *N.B.S. App Mathematics Series No. 14.*

[6] National Bureau of Standards, "Tables of the descending exponential," *N.B.S. App Mathematics Series No. 46.*

[7] National Bureau of Standards, "Tables of Normal Probability Functions," *N.I Applied Mathematics Series No. 23.*

[8] Z. W. BIRNBAUM AND FRED H. TINGEY, "One-sided confidence contours for probabi distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951) pp. 592–596.

ON THE INTEGRODIFFERENTIAL EQUATION OF TAKÁCS. I.

By Edgar Reich[1]

University of Minnesota

1. Introduction. This paper is devoted to a study of certain aspects of the mixed-type Markov process $\eta(t)$, originally treated by Takács [8]. It extends and unifies a number of results of previous workers.

Let $N(t)$, $N(0) = 0$, $t \geqq 0$ denote $\max \{n \mid t_n \leqq t\}$. We shall be especially interested in the case where $0 < t_1 < t_2 < \cdots$ are the events of an (in general) non-homogeneous Poisson process of density $\lambda(t) \geqq 0$. We assume that $\lambda(t)$ is Riemann integrable over all finite intervals (The homogeneous Poisson process corresponds to $\lambda(t) = \text{const.}$) Let $\chi_0, \chi_1, \chi_2, \cdots$ be a sequence of non-negative random variables. Except in a part of Section 5, they are mutually independent, and independent of $N(t)$; moreover, $H(x) = \Pr\{\chi_i \leqq x\}$ is the same for $i = 1, 2, \cdots$. Introducing the notations

$$\int_{-\infty}^{t} \chi(t)\, dN(t) = \chi_0 + \sum_{i=1}^{N(t)} \chi_i,\ L(x) = \begin{cases} 0, x \leqq 0 \\ 1, x > 0 \end{cases},$$

one may define (See Fig. 1)

$$\eta(t) = \int_{-\infty}^{t} \chi(u)\, dN(u) - \int_{0}^{t} L(\eta(u))\, du \tag{1.1}$$

It is sometimes instructive to formally redefine $\chi(t)$ as a stochastic process with $\chi(t)$, $\chi(t')$, $(t \neq t')$, independent, $\Pr\{\chi(t) \leqq x\} = H(x)$, $t > 0$. One then concludes immediately, from the functional form of (1.1) that $\eta(t)$ is a Markov process. Note that $\operatorname{var}(\eta(t + \Delta t) - \eta(t)) = O((\Delta t)^2)$, $t_i < t < t_{i+1}$, so that Feller's [5] function $a(t, x) = 0$.

In Section 2, the problem of finding the distribution of $\eta(t)$ will be reduced to finding the unique solution of a Volterra equation of the second kind. In Section 3, the corresponding result is found for the process $\eta^*(t)$, where, if t' is the first zero of $\eta(t)$,

$$\eta^*(t) = \begin{cases} \eta(t), t < t' \\ 0 \quad, t \geqq t'. \end{cases}$$

The work in Sections 2–4 generalizes results of Beneš [2] who treated the Takács process when $\lambda(t) = \text{const}$ (under somewhat milder restrictions on H). Section 5 contains some results on the asymptotic nature of $\eta(t)$, derived from a more general point of view than that employed in the preceding sections.

2. The Volterra equation for $\Pr\{\eta(t) = 0\}$.

Define $\Lambda(t) = \int_0^t \lambda(u)\, du$, $F(t,x) = \Pr\{\eta(t) \leqq x\}$, $F(t) = F(t,0) = \Pr$

Received November 4, 1957.

[1] A part of this work was done with support under Contract Nonr-710 (16).

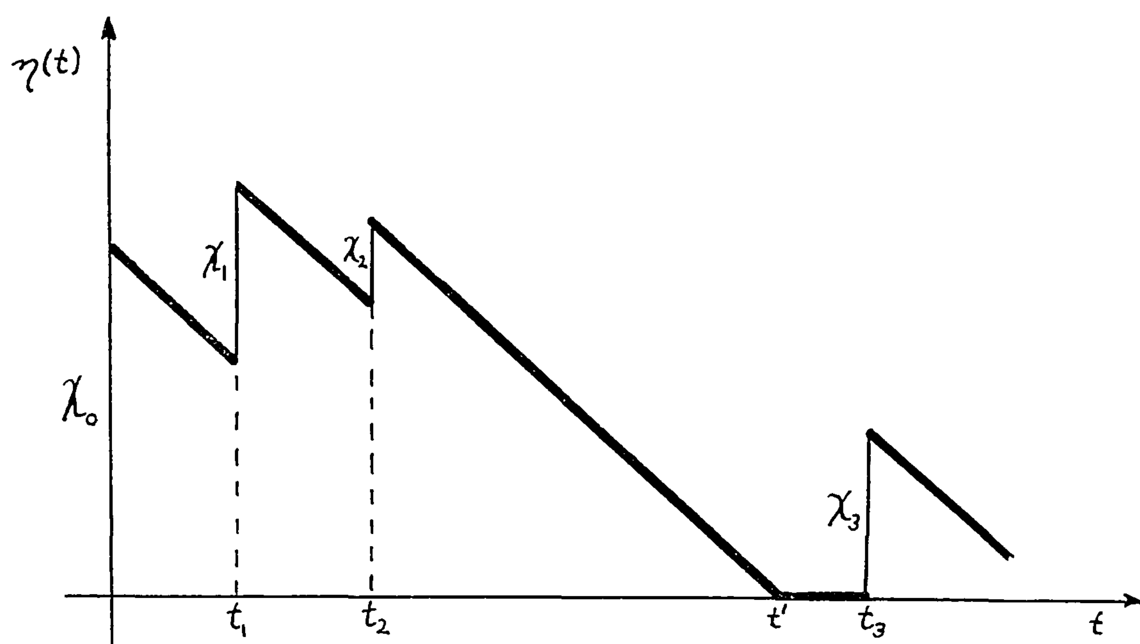


FIG. 1

$\{\eta(t) = 0\}$, $\psi(s) = Ee^{-s\chi_i}$, $i = 1, 2, \cdots$, $\Phi(t, s) = Ee^{-s\eta(t)}$, $\Phi(s) = \Phi(0, s) = Ee^{-s\chi_0}$, $(\Re s \geqq 0)$. It then follows [8] that $F(t, x)$ is continuous, $t \geqq 0$, $x > 0$, and[2]

$$\frac{\partial F(t, x)}{\partial t} = \frac{\partial F(t, x)}{\partial x} - \lambda(t)F(t, x) + \lambda(t) \int_{0-}^{x} H(x - y)\, d_y F(t, y).$$

Consequently,

$$\Phi(t, s) + s \int_0^t e^{s(t-u)-[1-\psi(s)][\Lambda(t)-\Lambda(u)]} F(u)\, du = \Phi(s)e^{st-[1-\psi(s)]\Lambda(t)}, \qquad \Re s \geqq 0. \tag{2.1}$$

Thus, if $F(t)$ is known, $F(t, x)$ can be computed by quadratures. Equation (2.1) contains two unknown functions, $F(t)$, and $\Phi(t, s)$, which might *a priori* lead one to believe that, unless the explicit relation between the two functions were brought into the picture, neither could be uniquely determined from (2.1) alone. However, by taking advantage of the regularity properties of Φ, (and certain additional regularity properties of λ, H) it turns out that (2.1) actually determines $F(t)$, and hence also $\Phi(t, s)$, uniquely. (Cf. Bailey [1] where regularity properties are used to solve a functional equation containing two unknown functions. See also [9], pp. 52–53.)

THEOREM 1. *Suppose* (i) $\lambda(t)$ ε $\mathfrak{L}^2$ *for every finite interval,* (ii) $H(x) = \int_0^x h(\xi)\, d\xi$, $e^{-cx}h(x)$ ε $\mathfrak{L}^2(0, \infty)$ *for some* $c \geqq 0$. *Then* $F(t)$ *is the unique continuous solution*

[2] Two functions of t will be written as equal, if they exist and are the same for almost all $t \geqq 0$.

of the Volterra equation of the second kind

$$g'(t) = F(t) + \int_0^t K(t, u)F(u)\, du, \quad \textit{where} \tag{2.2}$$

$$K(t, u) = \frac{1}{2\pi i} \frac{d}{dt} \text{P.V.} \int_{x-i\infty}^{x+i\infty} e^{(t-u)s - [\Lambda(t) - \Lambda(u)][1-\psi(s)]} \frac{ds}{s},$$

$$g'(t) = \frac{1}{2\pi i} \frac{d}{dt} \int_{x-i\infty}^{x+i\infty} \Phi(s) e^{ts - [1-\psi(s)]\Lambda(t)} \frac{ds}{s^2}, \qquad (x > c).$$

LEMMA 2.1. *If $x > 0$, $M > 0$, $-\infty < \gamma \leqq \gamma_0$, then $1/2\pi i \int_{x-iM}^{x+iM} e^{\gamma s}/s\, ds$ is bounded uniformly with respect to M, γ.*

PROOF. Let C_γ be the rectangular contour bounded by $s = x \pm iM$, and parallel lines extending to ∞ in the right (left) half plane if $\gamma < 0$ ($\gamma > 0$). Then

$$\frac{1}{2\pi i} \int_{C_\gamma} \frac{e^{\gamma s}}{s}\, ds = \frac{1 + \operatorname{sgn} \gamma}{2}.$$

$$\therefore \left| \frac{1}{2\pi i} \int_{x-iM}^{x+iM} \right| - 1 \leqq \left| \frac{1}{2\pi i} \int_{C_\gamma} - \frac{1}{2\pi i} \int_{x-iM}^{x+iM} \right|$$

$$\leqq \frac{K_1}{|\gamma| M} \leqq K_1, \quad \text{if} \quad |\gamma M| \geqq 1.$$

On the other hand, if $|\gamma M| \leqq 1$, then

$$\left| \frac{1}{2\pi i} \int_{x-iM}^{x+iM} \right| - \frac{1}{2} \leqq \left| \frac{1}{2\pi i} \int_{x-iM}^{x+iM} \frac{e^{\gamma s} - 1}{s}\, ds \right|$$

$$\leqq K_2 \int_{x-iM}^{x+iM} \frac{|\gamma s|}{|s|} |ds| = 2K_2 |\gamma| M \leqq 2K_2.$$

LEMMA 2.2. *If $x > 0$, $M > 0$, $0 \leqq \alpha \leqq \alpha_0$, $r(t) \in \mathfrak{L}^1(0, \infty)$, $R(s) = \int_0^\infty e^{-st} r(t)\, dt$, then* $1/2\pi i \int_{x-iM}^{x+iM} R(s) e^{\alpha s}\, ds/s$ is bounded uniformly with respect to M, α.

PROOF. Since the integral for $R(s)$ converges uniformly on the line $\Re s = x$,

$$L = \frac{1}{2\pi i} \int_{x-iM}^{x+iM} R(s) e^{\alpha s} \frac{ds}{s} = \int_0^\infty r(t) \left\{ \frac{1}{2\pi i} \int_{x-iM}^{x+iM} e^{(\alpha - t)s} \frac{ds}{s} \right\} dt.$$

Hence, by Lemma 2.1, $|L| \leqq K_3 \int_0^\infty |r(t)|\, dt$.

PROOF OF THEOREM 1. Dividing both sides of (2.1) by s^2, and integrating along the line $\Re s = x > c$ from $s = x - iM_1$ to $s = x + iM_2$, (M_1, $M_2 > 0$), we have

$$\frac{1}{2\pi i} \int_{x-iM_1}^{x+iM_2} \Phi(t, s) \frac{ds}{s^2} + \int_0^t \left\{ \frac{1}{2\pi i} \int_{x-iM_1}^{x+iM_2} e^{(t-u)s - [\Lambda(t) - \Lambda(u)][1-\psi(s)]} \frac{ds}{s} \right\} F(u)\, du$$

$$= \frac{1}{2\pi i} \int_{x-iM_1}^{x+iM_2} \Phi(s) e^{ts - \Lambda(t)[1-\psi(s)]} \frac{ds}{s^2}, \qquad t \geqq 0.$$

Since $\Phi(t, s)$ is regular, $|\Phi(t, s)| \leqq 1$, when $\Re s > 0$, the first integral on the left

side converges (absolutely) to zero as $M_1, M_2 \to \infty$, and the integral on the right side converges absolutely to the function

$$g(t) = \frac{1}{2\pi i} \int_{x-i\infty}^{x+i\infty} \Phi(s) e^{ts-\Lambda(t)[1-\psi(s)]} \frac{ds}{s^2}.$$

Hence

$$\text{(2.3)} \quad \lim_{M_1, M_2 \to \infty} \int_0^t \left\{ \frac{1}{2\pi i} \int_{x-iM_1}^{x+iM_2} e^{(t-u)s-[\Lambda(t)-\Lambda(u)][1-\psi(s)]} \frac{ds}{s} \right\} F(u)\, du$$

$$= g(t), \qquad t \geqq 0, x > c.$$

In particular, we shall henceforth take $M_1 = M_2 = M$, in order to make it possible to be able to invert the order of integration in (2.3). We can write

$$\frac{1}{2\pi i} \int_{x-iM}^{x+iM} e^{\alpha s-\beta[1-\psi(s)]} \frac{ds}{s} = \frac{1}{2\pi i} \int_{x-iM}^{x+iM} e^{\alpha s-\beta}[e^{\beta\psi(s)} - 1 - \beta\psi(s)] \frac{ds}{s}$$

$$+ \frac{e^{-\beta}}{2\pi i} \int_{x-iM}^{x+iM} e^{\alpha s} \frac{ds}{s} + \frac{\beta e^{-\beta}}{2\pi i} \int_{x-iM}^{x+iM} \psi(s) e^{\alpha s} \frac{ds}{s} = I + II + III$$

By the Riemann-Lebesgue Lemma, $\lim_{|y| \to \infty} |\psi(x + iy)| = 0$. By Parseval's equality,

$$\int_{-\infty}^{\infty} |\psi(x + iy)|^2\, dy = \int_0^{\infty} e^{-2cx}[h(x)]^2\, dx < \infty. \quad \text{Hence,}$$

$$|I| \leqq \beta K e^{\alpha x} \int_{x-i\infty}^{x+i\infty} |\psi(s)|^2 \frac{ds}{s} < \infty.$$

Therefore, as $M \to \infty$, I converges absolutely, and uniformly with respect to $\alpha, \beta, |\alpha| \leqq \alpha_0, |\beta| \leqq \beta_0$. By Lemma 2.1, $\lim_{M \to \infty} II = e^{-\beta}$, boundedly with respect $M > 0$, $0 \leqq \alpha \leqq \alpha_0$. By Lemma 2.2, $\lim_{M \to \infty} III = \beta e^{-\beta} H(\alpha)$, boundedly with respect to $M > 0$, $0 \leqq \alpha \leqq \alpha_0$. Thus we may rewrite (2.3) as the Volterra equation of the first kind,

$$\text{(2.4)} \quad \int_0^t G(t, u) F(u)\, du = g(t), \qquad t \geqq 0, \qquad (x > c), \qquad \text{where} \quad G(t, u)$$

$$= \text{P.V.} \frac{1}{2\pi i} \int_{x-i\infty}^{x+i\infty} e^{(t-u)s-[\Lambda(t)-\Lambda(u)][1-\psi(s)]} \frac{ds}{s} = \rho(\alpha, \beta) - \sigma(\alpha, \beta),$$

$$\text{(2.5)} \quad \rho(\alpha, \beta) = \frac{1}{2\pi i} \int_{x-i\infty}^{x+i\infty} e^{\alpha s-\beta}[e^{\beta\psi(s)} - 1 - \beta\psi(s)] \frac{ds}{s} + e^{-\beta},$$

$$\sigma(\alpha, \beta) = \beta e^{-\beta} H(\alpha), \qquad \alpha = t - u \geqq 0, \qquad \beta = \Lambda(t) - \Lambda(u).$$

Next we deal with the question of the existence and nature of the derivative $g'(t)$ for almost all $t \geqq 0$. First we focus on the existence and nature of

$$\frac{d}{dt} \int_0^t \rho(\alpha, \beta) F(u)\, du.$$

Since $\rho(\alpha, \beta)F(u)$ is continuous in t, u, $\rho(0, 0) = 1$,

$$\frac{d}{dt}\int_0^t \rho(\alpha, \beta)F(u)\,du = F(t) + \left[\frac{d}{dt}\int_0^\tau \rho(\alpha, \beta)F(u)\,du\right]_{\tau=t}$$

The partial derivatives ρ_α, ρ_β exist and are uniformly continuous in $0 \leqq \alpha \leqq \alpha_0$, $0 \leqq \beta \leqq \beta_0$. Thus $\Delta\rho = \rho_\alpha\Delta\alpha + \rho_\beta\Delta\beta + \epsilon_1\Delta\alpha + \epsilon_2\Delta\beta$, $\lim_{\Delta\alpha,\Delta\beta\to 0}\epsilon_i = 0$, ϵ_i uniformly bounded with respect to α, β, $0 \leqq \alpha \leqq \alpha_0$, $0 \leqq \beta \leqq \beta_0$. We have

$$\frac{\Delta\rho}{\Delta t} = \rho_\alpha + \epsilon_1 + \frac{(\rho_\beta + \epsilon_2)}{\Delta t}\int_t^{t+\Delta t} \lambda(v)\,dv.$$

Let $E = \{t \mid \Lambda'(t) = \lambda(t)\}$. We see that for $t \varepsilon E$, $\Delta\rho/\Delta t$ is bounded uniformly with respect to u, and $d\rho/dt = \rho_\alpha + \rho_\beta\lambda(t)$. Hence, by the bounded convergence theorem,

$$\frac{d}{dt}\int_0^\tau \rho F(u)\,du = \int_0^\tau \frac{\partial\rho}{\partial t}F(u)\,du,$$

and therefore

$$(2.6)\qquad \begin{aligned}\frac{d}{dt}\int_0^t \rho(\alpha, \beta)F(u)\,du &= F(t) + \int_0^t \frac{\partial\rho}{\partial t}F(u)\,du\\ &= \int_0^t \rho_\alpha F(u)\,du + \lambda(t)\int_0^t \rho_\beta F(u)\,du.\end{aligned}$$

We see, by (2.6), that

$$\frac{\partial\rho}{\partial t}\,\varepsilon\mathfrak{L}^2(\Delta), \qquad \Delta = \{(t, u) \mid 0 \leqq u \leqq t\}.$$

Also, $\int_0^t \partial\rho/\partial t\, F(u)\,du\ \varepsilon\ \mathfrak{L}^2$ (for every finite interval). Next, consider

$$\int_0^t \sigma(\alpha, \beta)F(u)\,du.$$

By noting that fact ([4], pp. 111 ff.) that for continuous $Q(u)$, and $h\ \varepsilon\ \mathfrak{L}^2$,

$$d/dt\int_0^t Q(u)H(t-u)\,du = \int_0^t Q(u)h(t-u)\,du = \text{continuous function of } t,$$

for all t, one finds that $d/dt\int_0^t \sigma(\alpha, \beta)F(u)\,du = \int_0^t \partial\sigma/\partial t\, F(u)\,du\ \varepsilon\ \mathfrak{L}^2$, with $\partial\sigma/\partial t\ \varepsilon\ \mathfrak{L}^2(\Delta)$. Thus (2.2) holds for almost all t, with

$$K(t, u) = \frac{\partial G(t, u)}{\partial t}\ \varepsilon\ \mathfrak{L}^2(\Delta),$$

and $g'(t)\ \varepsilon\ \mathfrak{L}^2$ over every finite interval. Under these conditions it is known [7] that (2.2) has a unique $\mathfrak{L}^2$ solution, $F(t)$; in particular, there is a unique continuous solution.

3. The Volterra equation for Pr $\{\eta^*(t) = 0\}$.

Define $B(t, x) = \Pr\{\eta^*(t) \leqq x\}$, $B(t) = \Pr\{\eta^*(t) = 0\} = \Pr\{t' \leqq t\}$, $\zeta(t, s) = Ee^{-s\eta^*(t)}$, $\Phi(s) = Ee^{-sx_0} = Ee^{-s\eta^*(0)}$. Then [2]

$$\frac{\partial B(t, x)}{\partial t} = \frac{\partial B(t, x)}{\partial x} - \lambda(t)B(t, x) + \lambda(t) \int_{0-}^{x} B(x - y, t)\, dH(y)$$
$$+ \lambda(t)[1 - H(x)]B(t).$$

Hence

$$\zeta(t, s) + \int_0^t \{s - \lambda(u)[1 - \psi(s)]\}e^{s(t-u)-[1-\psi(s)][\Lambda(t)-\Lambda(u)]}B(u)\, du$$
$$= \Phi(s)e^{st-[1-\psi(s)]\Lambda(t)}, \qquad \Re s \geqq 0. \tag{3.1}$$

THEOREM 2. *Under the same assumptions on* λ *and* H *as for Theorem* 1, $B(t)$ *is the unique continuous solution of the Volterra equation of the second kind*

$$g'(t) = B(t) + \int_0^t K^*(t, u)B(u)\, du, \tag{3.2}$$

where

$$K^*(t, u) = \frac{1}{2\pi i}\frac{d}{dt}\text{P.V.}\int_{x-i\infty}^{x+i\infty}$$
$$\cdot\{s - \lambda(u)[1 - \psi(s)]e^{s(t-u)-[1-\psi(s)][\Lambda(t)-\Lambda(u)]}\frac{ds}{s^2}, \qquad x > c.$$

PROOF. The proof proceeds as for Theorem 1, except that, before differentiation, the kernel now contains an additional term of the type

$$\rho^*(\alpha, \beta) = \frac{1}{2\pi i}\text{P.V.}\int_{x-i\infty}^{x+i\infty}[1 - \psi(s)]e^{\alpha s-\beta[1-\psi(s)]}\frac{ds}{s^2}$$
$$= \frac{1}{2\pi i}\int_{x-i\infty}^{x+i\infty} e^{\alpha s-\beta}\{[1 - \psi(s)][e^{\beta\psi(s)} - 1] - \beta\psi(s)\}\frac{ds}{s^2}$$
$$+ (\beta - 1)e^{-\beta}\int_0^\alpha H(\tau)\, d\tau.$$

This expression is treated in the same manner as $\rho(\alpha, \beta)$ was treated.

4. $\psi(s)$ **Regular at infinity.** We shall briefly remark on the practically important case when $\psi(s)$ is regular at infinity (e.g. when $\psi(s)$ is rational). This assumption regarding ψ is more restrictive than the assumptions in the hypotheses of Theorems 1 and 2, because by Pincherle's Theorem ([4], pg. 263),

$$\psi(s) = a_1 s^{-1} + a_2 s^{-2} + \cdots$$

is the Laplace transform of a density

$$h(t) = \frac{1}{2\pi i} = \int_\Gamma e^{ts}\psi(s)\, ds, \qquad t > 0,$$

where Γ is a contour, on and outside of which $\psi(s)$ is regular. In particular, if ψ is a rectangle on which $\Re s \leqq \delta > 0$, then we see that $|h(t)| \leqq Ke^{\delta t}$. Thus one may choose $c = 0$.

Instead of multiplying (2.1) (or (3.1)) by the "convergence factor" s^{-2}, it is sometimes more convenient to use $(s + \mu)^{-2}$, $\mu > 0$. For example, the kernel $G(t, u)$ of (2.4) then becomes

$$G(t, u) = \frac{1}{2\pi i} \int_\Gamma s(s + \mu)^{-2} e^{(t-u)s - [\Lambda(t) - \Lambda(u)][1-\psi(s)]} \, ds, \tag{4.1}$$

and $K(t, u) = (\partial/\partial t)G(t, u)$. For instance, if $\psi(s) = (1 + s)^{-1}$, and if we choose $\mu = 1$, then, if I_ν is the modified Bessel function,

$$G(t, u) = \begin{cases} e^{-(t-u)-[\Lambda(t)-\Lambda(u)]} I_0[2(t - u)^{1/2}(\Lambda(t) - \Lambda(u))^{1/2}] - (t - u)^{1/2} \\ (\Lambda(t) - \Lambda(u))^{-1/2} I_1[2(t - u)^{1/2}(\Lambda(t) - \Lambda(u))^{1/2}], \\ \text{if } \Lambda(t) \neq \Lambda(u), \\ e^{-(t-u)}[1 - (t - u)], \text{ if } \Lambda(t) = \Lambda(u). \end{cases}$$

This is rather similar to the kernel encountered by Clarke [3] by a completely different approach.

5. Asymptotic behavior of $\eta(t)$. Unless specifically stated, no restrictions regarding the distribution, or *independence* of the sequences $\{\chi_i\}$, $\chi_i \geqq 0$, and $\{t_n\}$, $0 < t_1 < t_2, \cdots$, shall be made in this section. Therefore, we cannot use the results of Sections 2–4, but must return to the fundamental relation (1.1).

LEMMA 5.1.

$$\eta(t) = \sup_{x>0} \left[\int_{t-x}^{t} \chi(u) \, dN(u) - x \right]$$

PROOF. Let $y = \{\max u \mid u \leqq t, \eta(u) = 0\}$. Then

$$\eta(t) = \int_y^t \chi(u) \, dN(u) - (t - y).$$

On the other hand,

$$\eta(t) = \eta(t - x) + \int_{t-x}^t \chi(u) \, dN(u) - \int_{t-x}^t L(\eta(u)) \, du \geqq \int_{t-x}^t \chi(u) \, dN(u) - x.$$

THEOREM 3. *If* $N(t) = \lambda t + o(t)$ *as* $t \to \infty$, $\sum_{i=1}^n \chi_i = \alpha n + o(n)$, *as* $n \to \infty$, $\lambda\alpha \leqq 1$, *then* $\eta(t) = o(t)$.

PROOF. We note first that the hypothesis implies that if $0 < \gamma_1 \leqq \gamma \leqq \gamma_2$, then

$$\lim_{t\to\infty} t^{-1} \left[\int_{-\infty}^{\gamma t} \chi(u) \, dN(u) - \gamma\alpha\lambda t \right] = 0,$$

uniformly with respect to γ. Let $\delta, \epsilon > 0$, be given. Then if $0 < x \leqq (1 - \delta)t$, there exists a $T_{\epsilon,\delta}$, such that

$$t^{-1} \left[\int_{t-x}^t \chi(u) \, dN(u) - x \right] = t^{-1} \left[\int_{-\infty}^t \chi(u) \, dN(u) - \int_{-\infty}^{(1-x/t)} \chi(u) \, dN(u) \right]$$
$$- (x/t) \leqq \alpha\lambda - (1 - x/t)\alpha\lambda + \epsilon - x/t \leqq \epsilon,$$

if $t > T_{\epsilon,\delta}$. On the other hand, if $x > (1-\delta)t$, there exists a T_ϵ such that

$$t^{-1}\left[\int_{t-x}^{t} \chi(u)\, dN(u) - x\right]$$
$$\leq t^{-1}\left[\int_{-\infty}^{t} \chi(u)\, dN(u)\right] - x/t \leq \alpha\lambda + \epsilon - (1-\delta) \leq \epsilon + \delta,$$

if $t > T_\epsilon$. Hence $\eta(t)\,/\,t \leq \epsilon + \delta$ if $t > \max\,(T_\epsilon\,,\ T_{\epsilon,\delta})$.

Corollary. *If $N(t)$ is a Poisson process with cumulative mass $\Lambda(t) = \lambda t + o(t)$ as $t \to \infty$, $\sum_{i=1}^{n} \chi_i = cn + o(n)$, as $n \to \infty$, $\lambda\alpha \leq 1$, then $\eta(t) = o(t)$* with probability 1.

Proof. If $\Lambda(\infty) < \infty$, the result is trivial, as then $\eta(t) = O(1)$, with probability one. Assume $\Lambda(\infty) = \infty$. Let $N^*(t, \omega)$, $\omega \in \Omega$, be a homogeneous Poisson process with unit density. Then $N(t, \omega) = N^*(\Lambda(t), \omega)$ is a Poisson process with density $\lambda(t)$. Hence

$$\lim_{t\to\infty} \frac{N(t,\omega)}{t} = \lim_{t\to\infty} \frac{N^*(\Lambda(t),\omega)}{\Lambda(t)} \frac{\Lambda(t)}{t} = \lambda, \qquad \text{a.a.}\omega.$$

The following result follows from results of Kiefer and Wolfowitz [6], after some elementary transformations.

Theorem 4. *Suppose $\Lambda(t) = \lambda t + O(1)$, as $t \to \infty$. If $\{\chi_i\}$ are independent of each other and $N(t)$, and are equidistributed, and if $\overline{\chi_i}\lambda < 1$, $\overline{\chi_i^2} < \infty$, then*

$$E\eta(t_n + 0) = O(1),$$

as $n \to \infty$.

The hypothesis on $\Lambda(t)$ is satisfied, e.g., if $\lambda(t)$ is periodic with mean λ. It may be shown, by counterexample, that the conclusion of Theorem 4 becomes false if the hypothesis is weakened to $\Lambda(t) = \lambda t + o(t)$.

REFERENCES

[1] N. T. J. Bailey, "A continuous time treatment of a simple queue using generating functions," *Roy. Stat. Soc. B*, Vol. 16 (1954), 288–291.

[2] V. E. Beneš, "On queues with Poisson arrivals," *Ann. Math. Stat.*, Vol. 28 (1957), 670–677.

[3] A. B. Clarke, "A waiting line process of Markov type," *Ann. Math. Stat.*, Vol. 27 (1956), 452–459.

[4] G. Doetsch, *Handbuch der Laplace-Transformation I*, Birkhäuser Verlag, 1950.

[5] W. Feller, "Zur Theorie der stochastischen Prozesse," *Math. Ann.*, Vol. 113 (1937), 113–160.

[6] J. Kiefer, and J. Wolfowitz, "On the Theory of Queues with Many Servers," *Trans. Amer. Math. Soc.*, Vol. 78 (1955), 1–18.

[7] F. Smithies, "On the theory of linear integral equations," *Proc. Cambridge Philos. Soc.*, Vol. 31 (1935), 76–84.

[8] L. Takács, "Investigation of waiting time problems by reduction to Markov processes," *Acta Math. Hung.*, Vol. 6 (1955), 101–129.

[9] G. Doetsch, *Handbuch der Laplace-Transformation III*, Birkhäuser Verlag, 1956.

NOTES

PROBABILITIES OF HYPOTHESES AND INFORMATION-STATISTICS IN SAMPLING FROM EXPONENTIAL-CLASS POPULATIONS

By Morton Kupperman

The George Washington University

1. Summary. This paper is concerned with inequalities connecting probabilities of hypotheses using Bayes' theorem (a posteriori probabilities), a priori probabilities, and Kullback-Leibler information-statistics in sampling from populations belonging to the exponential class of populations. As a corollary, it is shown that if it is known that the a priori probabilities are all equal, the choice of the hypothesis with the minimum Kullback-Leibler information-statistic is the same as the choice of the hypothesis with the maximum a posteriori probability, and conversely.

2. Introduction. Suppose that an event E can occur only if one of the set of r exhaustive and incompatible (mutually exclusive) events $H_1, H_2, \cdots, H_r$ occurs. The a priori probabilities of these latter events (which we may call hypotheses) are denoted by $\alpha_1, \alpha_2, \cdots, \alpha_r$ respectively, where $\alpha_m > 0$ and $\sum_{m=1}^{r} \alpha_m = 1$. The conditional probabilities for E to occur, assuming the occurrence of H_m, are denoted by $p(E \mid H_m)$, $m = 1, 2, \cdots, r$. The a posteriori probabilities of H_m, given that E has occurred, are denoted by $p(H_m \mid E)$. Bayes' theorem (see, for example, Uspensky [16]) states that

$$p(H_m \mid E) = \frac{\alpha_m p(E \mid H_m)}{\sum_{j=1}^{r} \alpha_j p(E \mid H_j)}, \qquad \text{for } m = 1, 2, \cdots, r.$$

A discrete multivariate and multiparameter population will be said to belong to the exponential class of populations (cf. Blackwell and Girshick [1] and Girshick and Savage [5]) if its probability distribution can be represented by

$$p(\mathbf{x}, \theta) = q(\theta)\, r(\mathbf{x}) \exp\left\{\sum_{i=1}^{h} s_i(\theta)\, t_i(\mathbf{x})\right\},$$

where $\mathbf{x}$ is the row vector $\mathbf{x} = (x_1, x_2, \cdots, x_k)$, θ is the row vector $\theta = (\theta_1, \theta_2, \cdots, \theta_h)$, $q(\theta)$ and $r(\mathbf{x})$ are nonnegative functions of θ and $\mathbf{x}$ respectively, and the parameter space is assumed to be an open convex set in an h-dimensional Euclidean space. We have k variates and h parameters, with the number of products in the exponent of e being h. Examples of discrete populations of the exponential class are the binomial distribution with the single parameter p,

Received September 25, 1956; revised November 2, 1957.

the Poisson distribution, the geometric distribution, the multinomial distribution, and the multivariate Poisson distribution.

Now consider r populations of the exponential class, each of the same functional form but differing only in their parameters. Let the probabilities be given by $p(\mathrm{x}, \theta_m) > 0$, where $\sum \mathrm{x} p(\mathrm{x}, \theta_m) = 1$ for $m = 1, 2, \cdots, r$. Suppose that we have a single random sample of N independent observations from one of these r populations (we do not know which population) and we wish to decide, on the basis of the sample values, which of the r populations is the most likely source of the sample. We shall use the term "most likely" in the sense of "having the largest a posteriori probability" and we shall assume that the a priori probabilities α_m are already known.

Let E denote the random sample and let $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_h)$ denote the maximum-likelihood estimate of θ.

3. Inequalities. The information measure $I(1:2)$ was introduced by Kullback and Leibler [12] as a generalization to the abstract case of a definition of information independently introduced in 1948 by Shannon [15] and by Wiener [17]. (See also Kullback [9], [10], and [11] for uses in statistics of $I(1:2)$. $I(1:2)$ has recently been termed "Kullback-Leibler information number" (Chernoff [3]) and "K-L information number" (Bradt and Karlin [2]).)

We obtain for two discrete populations of the exponential class

$$I(1:2) = \sum_{\mathrm{x}} p(\mathrm{x}, \theta_1) \log \frac{p(\mathrm{x}, \theta_1)}{p(\mathrm{x}, \theta_2)}$$

$$= \log \frac{q(\theta_1)}{q(\theta_2)} + \sum_{i=1}^{h} \left[\left\{ s_i(\theta_1) - s_i(\theta_2) \right\} \cdot E_1 \left\{ t_i(\mathrm{x}) \right\} \right],$$

where the probabilities for the first population are given by $p(\mathrm{x}, \theta_1)$, the probabilities for the second population are given by $p(\mathrm{x}, \theta_2)$, and E_1 denotes expected values with respect to the first population. The logarithms are natural logarithms.

We now define the Kullback-Leibler information-statistic for a random sample of N independent observations from the mth population as

$$\hat{I}_m = N \sum_{\mathrm{x}} p(\mathrm{x}, \hat{\theta}) \log \frac{p(\mathrm{x}, \hat{\theta})}{p(\mathrm{x}, \theta_m)}$$

$$= N \log \frac{q(\hat{\theta})}{q(\theta_m)} + N \sum_{i=1}^{h} \left[\left\{ s_i(\hat{\theta}) - s_i(\theta_m) \right\} \cdot \left(E \left\{ t_i(\mathrm{x}) \right\} \right)_{\theta=\hat{\theta}} \right].$$

In $I(1:2)$, which is a functional of the vectors θ_1 and θ_2 only, θ_1 has been replaced by the maximum-likelihood estimate $\hat{\theta}$ and θ_2 has been replaced by the set of parameters θ_m of the hypothetical mth population. The sum has been multiplied by N since the information measure for N independent observations is N times the information measure for a single observation.

The Kullback-Leibler information-statistic for samples from discrete popula-

tions of the exponential class (as well as for samples from more general statistical populations, discrete or continuous, univariate or multivariate, uniparameter or multiparameter) has useful applications in mathematical statistics. If we set up a null hypothesis that the given sample of N independent observations was randomly drawn from the specified mth population, then it can be shown that $2\hat{I}_m$ as defined above is asymptotically distributed as chi-square with h degrees of freedom when the null hypothesis is true (Kupperman [13], [14]).

We shall now show that the following inequality relationships exist connecting the a posteriori probabilities, the a priori probabilities, and the Kullback-Leibler information-statistics:

THEOREM. *For two discrete populations H_m and H_n of the exponential class we have $p(H_m \mid E) \geqq p(H_n \mid E)$ if and only if $\hat{I}_m \leqq \hat{I}_n + \log(\alpha_m/\alpha_n)$, with both relations being equalities or strict inequalities simultaneously.*

PROOF. From $p(H_m \mid E) \geqq p(H_n \mid E)$ we obtain, using Bayes' theorem and simplifying,

$$[q(\theta_m)]^{-N} \exp\left\{-\sum_{j=1}^{N}\sum_{i=1}^{h} s_i(\theta_m)\, t_i(X_j)\right\}$$

$$\leqq [q(\theta_n)]^{-N} \exp\left\{-\sum_{j=1}^{N}\sum_{i=1}^{h} s_i(\theta_n)\, t_i(X_j)\right\} \cdot \frac{\alpha_m}{\alpha_n},$$

where X_j is the value of the jth observation on x, $j = 1, 2, \cdots, N$. Now it can be shown (Kupperman [14]) that for populations of this class we have identically

$$\left(E\left\{t_i(x)\right\}\right)_{\theta=\hat{\theta}} = \frac{1}{N}\sum_{j=1}^{N} t_i(X_j).$$

(The discrete populations of the exponential class now being considered belong to the class of distributions admitting sufficient estimates of the parameters θ; these are distributions of the Koopman-Pitman type.) Hence by multiplying both sides of the inequality by the positive quantity

$$[q(\hat{\theta})]^{N} \exp\left\{N \sum_{i=1}^{h} s_i(\hat{\theta}) \cdot \left(E\left\{t_i(x)\right\}\right)_{\theta=\hat{\theta}}\right\}$$

and taking logarithms, we obtain

$$\hat{I}_m \leqq \hat{I}_n + \log\frac{\alpha_m}{\alpha_n}.$$

Since the steps of the proof are all reversible, the theorem is proved. The following corollary is an immediate consequence of this theorem:

COROLLARY. *If the a priori probabilities are all equal, the choice of the hypothesis (or hypotheses) with the minimum Kullback-Leibler information-statistic is the same as the choice of the hypothesis (or hypotheses) with the maximum a posteriori probability, and conversely.*

4. Continuous exponential-class populations. Although the preceding two

sections are concerned with discrete populations, it may be remarked that the theorem and corollary may easily be extended to samples from continuous populations of the exponential class (such as the univariate normal distribution, the chi-square distribution, and the multivariate normal distribution with k means and $k(k + 1)/2$ parameters in the variance-covariance matrix). We make use of Bayes' theorem for continuous distributions (Kolmogorov [8], p. 46), use the likelihood of the observed sample instead of the probability of the observed sample, and follow the same steps as in the proof in the discrete case. The statements concerning $(E\{t_i(x)\})_{\theta=\hat{\theta}}$ and the asymptotic distribution of $2\hat{I}_m$ remain valid for continuous as well as discrete distributions of the exponential class.

5. Application. The theorem and the corollary are applicable to problems in which the a priori probabilities can be expressed in exact numerical form and thus the application of Bayes' theorem is legitimate, as, for example, in Mendelian hypotheses (see David [4], Chapter VIII).

In connection with the theorem and corollary, it may be remarked that the statements hold true if common logarithms (or logarithms to any base) are used in place of natural logarithms. This point is of importance, for in practical work common logarithms are more frequently used. However, in connection with the approximation of the large-sample distribution of $2\hat{I}$ by a chi-square distribution, it is important that natural logarithms be used, or that if common logarithms have been used $2\hat{I}$ be multiplied by $\log_e 10$, or 2.30259 approximately.

In conclusion, it may be remarked that if we were to use the corollary and decide always to accept the hypothesis for which $\hat{I}$ is the minimum without regard to the a priori probabilities involved, then we are in effect tacitly assuming that the a priori probabilities are equal, which is Bayes' postulate (as distinguished from Bayes' theorem).

The connection between information theory and inverse probability has been noted by Good [7], who is also concerned with the terminology and notation of information theory, particularly as it is applicable to communication theory. Reference should also be made to Good [6] for an informative discussion on Bayes' theorem and inverse probability.

6. Acknowledgment. The author wishes to thank Professor S. Kullback and the referees for suggesting the generalization incorporated in the results of this paper, which were derived originally for the special case of multinomial sampling.

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley & Sons, Inc., New York, 1954.

[2] R. N. Bradt and S. Karlin, "On the design and comparison of certain dichotomous experiments," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 390–409.

[3] H. Chernoff, "Large-sample theory: Parametric case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1–22.

[4] F. N. David, *Probability Theory for Statistical Methods*, Cambridge University Press, Cambridge, 1949.

[5] M. A. Girshick and L. J. Savage, "Bayes and minimax estimates for quadratic loss

functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley and Los Angeles, 1951, pp. 53–73.
[6] I. J. Good, *Probability and the Weighing of Evidence*, Charles Griffin & Company Limited, London, 1950.
[7] I J. Good, "Some terminology and notation in information theory," *Monograph No 155R*, The Institution of Electrical Engineers, London, 1955.
[8] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, 1933.
[9] S Kullback, "An application of information theory to multivariate analysis," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 88–102.
[10] S. Kullback, "Certain inequalities in information theory and the Cramér-Rao inequality," *Ann. Math. Stat.*, Vol. 25 (1954), pp 745–751
[11] S Kullback, "An application of information theory to multivariate analysis, II," *Ann. Math. Stat*, Vol. 27 (1956), pp. 122–146
[12] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math Stat*, Vol. 22 (1951), pp. 79–86.
[13] M Kupperman, "Further applications of information theory to multivariate analysis and statistical inference, (preliminary report)," (abstract), *Ann. Math Stat*, Vol. 27 (1956), p. 1184
[14] M Kupperman, "Further applications of information theory to multivariate analysis and statistical inference," Ph D dissertation on file at the library of The George Washington University, February 1957
[15] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech J*, Vol. 27 (1948), pp 379–423 and 623–656
[16] J V Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill Book Company, Inc, New York, 1937.
[17] N. Wiener, *Cybernetics*, John Wiley & Sons, Inc, New York, 1948

ON THE DISTRIBUTION OF 2 × 2 RANDOM NORMAL DETERMINANTS[1]

By W L. Nicholson[2]

Princeton University

1. Summary. The c.d.f. of a 2 × 2 random determinant with mutually independent normally distributed entries is derived as an infinite series. Error functions that bound the tail of this series facilitate numerical calculation. Conditions are imposed on four variable quadratic forms for this distribution to apply. A normal approximation to the distribution is suggested.

2. Introduction. Let X_1, X_2, X_3 and X_4 be mutually independent random variables, each normally distributed, with means μ_1, μ_2, μ_3 and μ_4, and common variance σ^2. Let D be the random determinant,

$$D = \begin{vmatrix} X_1 X_2 \\ X_3 X_4 \end{vmatrix} = X_1 X_4 - X_2 X_3.$$

Received June 18, 1956, revised December 9, 1957

[1] Work sponsored by the Office of Naval Research

[2] Now employed by the General Electric Company, Richland, Washington.

If all the μ_i vanish the p.d.f. of D/σ^2 is easily calculated to be the Laplace distribution [5],

$$\tfrac{1}{4} \exp \{-\tfrac{1}{2}| x |\}.$$

When the μ_i are not zero the distribution is, in general, skewed and not expressible in a simple closed form.

Craig [1] derived the p.d.f. of the product of two normal variables (not necessarily independent) as an infinite series of Bessel functions. Theoretically, his result plus the convolution formula for density functions determines the p.d.f. of D. However, the form of such an answer is not particularly adapted to numerical work. Most methods for handling the distribution problems connected with normal variable quadratic forms are not applicable here. The reasons for this are, first, that D is not a definite form, and, second, that it cannot be represented as a linear combination of central Chi-Square variables. The former obstacle can be overcome to a measured degree by several different procedures; e.g., Pitman's and Robbins' method of mixtures [6] and Gurland's Laguerrian expansions [3]. The latter causes more difficulty. There does not seem to be an adequate technique available to handle linear combinations of non-central Chi-Square variables.

Our approach is basically a brute force method consisting of straightforward inversion of the characteristic function of D. The independence and homoscedasticity assumptions cannot be relaxed without greatly complicating this inversion problem. In the process a single integration leads to the c.d.f. of D. Percentage points are immediately available without resorting to quadratures.

In the sequel $\sigma^2 = 1$. There is no loss of generality in this simplification, since σ^2 appears as a scale parameter in the distribution of D; i.e., we derive the distribution of the normalized variable D/σ^2.

The characteristic function of D is easily calculated to be

$$\phi_D(t) = E(e^{itD}) = (1 + t^2)^{-1} \exp\left\{\frac{-\Lambda t^2 + 2i\Delta t}{2(1 + t^2)}\right\} \tag{2.1}$$

where

$$\Lambda = \mu_1^2 + \mu_2^2 + \mu_3^2 + \mu_4^2, \qquad 0 \leqq \Lambda < +\infty,$$

$$\Delta = \begin{vmatrix} \mu_1 & \mu_2 \\ \mu_3 & \mu_4 \end{vmatrix} = \mu_1 \mu_4 - \mu_2 \mu_3, \qquad -\frac{\Lambda}{2} \leqq \Delta \leqq \frac{\Lambda}{2}. \tag{2.2}$$

Thus, we see that the distribution of D depends on the means only in the form Λ and Δ. Expand $\log \phi_D(t)$ in powers of t to get the semi-invariants of D as

$$\alpha_{2k} = (2k)!\left(\frac{\Lambda}{2} - \frac{1}{k}\right) \qquad k = 1, 2, \cdots;$$

$$\alpha_{2k+1} = (2k + 1)!\,\Delta \qquad k = 0, 1, 2, \cdots. \tag{2.3}$$

The mean and variance of D are

$$\mu_D = \alpha_1 = \Delta, \qquad \sigma_D^2 = \alpha_2 = \Lambda + 2. \tag{2.4}$$

The coefficients of skewness and excess are

$$\gamma_1 = \frac{\alpha_3}{\alpha_2^{3/2}} = \frac{6\Delta}{(\Lambda + 2)^{3/2}}, \tag{2.5}$$
$$\gamma_2 = \frac{\alpha_4}{\alpha_2^2} = \frac{12(\Lambda + 1)}{(\Lambda + 2)^2}.$$

The distribution is skewed if and only if $\Delta \neq 0$. From (2.5) this skewness is never great. In fact $|\gamma_1| \leqq 2\sqrt{6}/3$ and $\gamma_1 = 0(\Lambda^{1/2})$ for large Λ. The excess γ_2 is monotone decreasing in Λ with a maximum value of three for $\Lambda = 0$. Also, $\gamma_2 = 0(\Lambda)$ for large Λ. Thus, if at least one μ_i is large the distribution is almost symmetric and of approximately the same peakedness as that of the normal. In Section 4 we show that D (appropriately normalized) is approximately normally distributed for large Λ.

3. Exact Distribution. The functional form of the characteristic function (2.1) indicates that the p.d.f., $f_{\Lambda,\Delta}$, and the c.d.f., $F_{\Lambda,\Delta}$, of D satisfy for all real x

$$f_{\Lambda,\Delta}(x) = f_{\Lambda,-\Delta}(-x), \qquad F_{\Lambda,\Delta}(x) = 1 - F_{\Lambda,-\Delta}(-x). \tag{3.1}$$

Hence we need only consider the distribution of D for negative argument. In the remainder of the paper x always satisfies $x \leqq 0$. The c.d.f. of D is not expressible in a simple closed form (unless $\Delta = \Lambda/2$.). Introduction of an appropriate error term does make it possible to represent it as a damped polynomial in $|x|$ with coefficients that are elementary functions of Λ and Δ. Let R be any set of non-negative integers. The c.d.f of D can be written as (see Sec. 5)

$$F_{\Lambda,\Delta}(x) = \sum_{r \in R} \sum_{t=0}^{r} h(r,t) g(r, t \mid \Delta, \Lambda/2, |x|) + L, \tag{3.2}$$

where L satisfies

$$0 \leqq L < \frac{1}{2} \sum_{r \in R} e^{-\Lambda/2} \frac{\left(\frac{\Lambda}{2}\right)^r}{r!} \qquad \Delta \geqq 0, \tag{3.3}$$
$$0 \leqq L < \frac{1}{2} \sum_{r \in R} e^{-\Lambda/2} \frac{\left(\frac{\Delta + |\Delta|}{2}\right)^r}{r!} \qquad \Delta < 0.$$

The auxiliary functions h and g are defined by

$$h(r,t) = \sum_{j=0}^{r-t} \binom{2r - t + 1}{j} \left(\frac{1}{2}\right)^{2r - t + 1}, \tag{3.4}$$
$$g(r, t \mid a, b, c) = e^{-(b+c)} \sum_{j=0}^{t} \frac{a^j (b - a)^{r-j} c^{t-j}}{j!(r - j)!(t - j)!}.$$

Here $h(r, t)$ is just the probability of not more than $r - t$ heads on $2r - t + 1$ flips of an unbiased coin. Several tables of h are available; e.g., [7]. The function

g satisfies a number of recursion formulae. The most useful of these for computational purposes is

(3.5) $$tg(r, t \mid a, b, c) = ag(r-1, t-1 \mid a, b, c) + cg(r, t-1 \mid a, b, c).$$

The boundary condition,

(3.6) $$g(r, 0 \mid a, b, c) = e^{-(a+c)} \left\{ e^{-(b-a)} \frac{(b-a)^r}{r!} \right\},$$

and (3.5) provide a rapid method of generating a matrix of g values for any triple $(a, b, c) = (\Delta, \Lambda/2, |x|)$. The right side of (3.6) is most easily calculated as the product of two tabular values. The bracket term as a Poisson density is tabled (see [4]).

The bound (3.3) on L is quite good for $\Delta \geqq 0$. Numerical checks show, for example, that for values of Λ of the order of ten a bound of 0.01 on L adds only one integer to R over and above that necessary to give an error $\leqq 0.01$. For $\Delta < 0$ the bound is admittedly rather poor and certainly could be improved.

To minimize the calculation necessary to evaluate $F_{\Lambda,\Delta}(x)$ the set R should contain as few elements as possible. To accomplish this and still maintain a specified bound on the error R should consist only of integers in an appropriate interval including $\Lambda/2$ (at least when $\Delta \geqq 0$). However, from the standpoint of iterated computation of the g function the optimum R set is $\{0, 1, 2, \cdots, M\}$ for suitable M. In this case the bound (3.3) on L is, except possibly for an exponential factor, the tail area of a Poisson distribution; its value can be read directly from tables [4].

At least three values of Δ lead to extreme simplification in the formula (3.2). These values are the maximum and the minimum Δ value for fixed Λ, and $\Delta = 0$. The simplified forms make possible several quickly computed bounds on $F_{\Lambda,\Delta}(x)$. The simplifications are

(3.7)
$$F_{\Lambda,-\Lambda/2}(x) = e^{-(\Lambda/2)-|x|} \sum_{r \in R} \sum_{t=0}^{r} \frac{2^{r-t+1}-1}{2^{r-t+1}} \frac{|x|^t}{t!} \frac{(\Lambda/2)^r}{r!} + L;$$
$$F_{\Lambda,0}(x) = e^{-(\Lambda/2)-|x|} \sum_{r \in R} \sum_{t=0}^{r} h(r, t) \frac{|x|^t}{t!} \frac{(\Lambda/2)^r}{r!} + L;$$
$$F_{\Lambda,\Lambda/2}(x) = \tfrac{1}{2} e^{-(\Lambda/4)-|x|}.$$

The bound (3.3) on L (with $\Delta \geqq 0$) is also applicable for the first two lines of (3.7). Since $F_{\Lambda,\Delta}(x)$ for fixed Λ and x is a monotone decreasing function of $\Delta(-\Lambda/2 \leqq \Delta \leqq \Lambda/2)$, the following inequalities are immediately available.

(3.8)
$$F_{\Lambda,-\Lambda/2}(x) \geqq F_{\Lambda,\Delta}(x) \geqq F_{\Lambda,0}(x) \qquad \Delta \leqq 0,$$
$$F_{\Lambda,0}(x) \geqq F_{\Lambda,\Delta}(x) \geqq F_{\Lambda,\Lambda/2}(x) \qquad \Delta \geqq 0.$$

A simple but interesting application of (3.8) is the following bound on the probability that D is negative.

$$\frac{1}{2}e^{-\Delta/4} \leqq \Pr\{D \leqq 0\} \leqq \frac{1}{2} \qquad \Delta \geqq 0,$$
(3.9)
$$\frac{1}{2} \leqq \Pr\{D \leqq 0\} \leqq 1 - \frac{1}{2}e^{\Delta/4} \qquad \Delta \leqq 0.$$

These are the best bounds possible that are functions of Δ only.

The quadratic form D is by no means representative of the class of all such forms in four variables. In general, the distribution of a four variable form is more complicated than that of D. There are certain special cases, however, when the distribution function (3 2) does apply. The following theorem gives a necessary and sufficient condition on four normal variables and on the accompanying form for the distribution to be that of this paper.

THEOREM. *Let $X = (X_1, X_2, X_3, X_4)$ be a random vector distributed by a multivariate normal law with mean vector μ and covariance matrix Σ. Let M be a 4×4 symmetric matrix. The quadratic form $XMX'/2d$ is distributed according to the law* (3.2) *if and only if the eigenvalues of the matrix $M\Sigma$ are d, d, $-d$ and $-d$. If such is the case $\Lambda = \mu\,\Sigma^{-1}\mu'$ and $\Delta = \mu M\mu'/2d$.*

A proof is easily constructed by identifying the characteristic function of $XMX'/2d$ with (2.1).

4. Normal Approximation. Let $\tilde{D} = D/(2 + \Lambda)^{\frac{1}{2}}$. Then, if Λ increases without bound in such a manner that $\Delta/(2 + \Lambda)^{\frac{1}{2}} \to a$, we have from (2.1) that $\phi_{\tilde{D}}(t) \to \exp(iat - t^2/2)$. So, by the continuity theorem for characteristic functions [2], for large Λ $\tilde{D}$ is approximately normal with mean $\Delta/(2 + \Lambda)^{\frac{1}{2}}$ and unit variance. The question of how large Λ must be for the approximation to render reasonable accuracy is quite difficult to answer. The following remarks are offered to give some insight into this problem. Clearly, the rapidity of the convergence depends upon Δ and $|x|$ in some fashion For $\Delta = 0$ the approximation is very good since this is the symmetric case. With Λ fixed the accuracy decreases as $|x|$ increases. Numerical checks indicate that for $|x|$ less than three and Λ about 20 the relative error in the c. d. f. is less than 5%. For the general case with Δ not too far different from zero the accuracy seems to be roughly a monotone decreasing function of $|x| + \Delta$ for fixed Λ. With $|x| + \Delta$ less than four and Λ about 20 the relative error is less than 7%. For large numerical values of Δ the approximation is extremely poor. For example, if $\Delta = 0(\Lambda)$ and if $|x| > 0$, then the relative error approaches 100% as Λ increases.

5. Derivation of Exact Distribution of D. Since ϕ_D is Lebesque integrable the Lévy inversion formula [2] gives the p.d.f. of D as

(5.1)
$$f_{\Lambda,\Delta}(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixt}\phi_D(t)\,dt = \frac{e^{-\Lambda/2}}{2}\sum_{j,k=0}^{\infty}\frac{\Delta^j(\Lambda/2)^k}{j!k!(j+k)!} \times \frac{\partial^{2j+k}}{\partial x^j\partial z^{j+k}}Q(x,z)\bigg|_{z=1}^{x=x}.$$

Here,

(5.2)
$$Q(x,z) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{e^{-ixt}}{z+t^2}\,dt = z^{-1/2}e^{-|x|z^{1/2}},$$

for $x \leqq 0$ and $z > 0$.

Evaluate the jth partial derivative of Q with respect to x first and integrate $f_{\Lambda,\Delta}(x)$ over the interval $(-\infty, x)$. Then (5.1) becomes

$$(5.3) \qquad F_{\Lambda,\Delta}(x) = \frac{e^{-\Lambda/2}}{2} \sum_{r \epsilon R} \left(\frac{1}{r!}\right)^2 \frac{\partial^r}{\partial z^r} \left[z^{-1} \left(\Delta z^{\frac{1}{2}} - \frac{\Lambda}{2}\right)^r e^{-|x| z^{1/2}} \right|_{z=1} + L,$$

where R is any set of non-negative integers and L consists of the remainder of the series; i.e., those terms such that $r \not\epsilon R$ (here $r = j + k$).

Rewrite the factor $(\Delta z^{1/2} - \Lambda/2)^r$ as an rth partial derivative of the appropriate exponential function of W evaluated at zero. Use Leibnitz's Rule for differentiating a product to compute the rth partial with respect to z of the resulting function after choosing one product factor as $z^{-1/2}$. Employ the identity,

$$(5.4) \qquad \frac{\partial^P}{\partial z^P} z^{-\frac{1}{2}} e^{-a z^{1/2}} \bigg|_{z=1} = (-\tfrac{1}{2})^P P! \, e^{-a} \sum_{i=0}^{P} \binom{2P - i}{P} \frac{(2a)^i}{i!},$$

with $a = |x| - \Delta W$, and complete the differentiation with respect to W. Considerable algebraic simplification involving routine summing of finite combinatorial type series gives the form (3.2) after the appropriate identifications have been made with the h and g functions defined by (3.4). The bounds (3.3) and the simplifications (3.7) result from straightforward algebra. Details are omitted.

6. Acknowledgement. The author expresses his appreciation to Dr. L. Marcus whose interest in this problem gave impetus for the paper. Thanks are also due to Professor J. W. Tukey for several helpful suggestions.

REFERENCES

[1] C. C. Craig, "On the frequency function of xy," *Ann. Math. Stat.*, Vol. 7 (1936), pp 1–15.

[2] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, Inc., New York, 1953.

[3] J. Gurland, "Distribution of quadratic forms and ratios of quadratic forms," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 416–427.

[4] E. C. Molina, *Poisson's Exponential Binomial Limit*, D. Van Nostrand Company, Inc., New York, 1942.

[5] H. Nyguist, S. O. Rice and J. Riordan, "The distribution of random determinants," *Quart. Appl. Math.*, Vol. 42 (1954), pp. 97–104.

[6] H. Robbins and E. J. G. Pitman, "Application of the method of mixtures to quadratic forms in normal variables," *Ann. Math. Stat.*, Vol. 20 (1949), ppl 552–560.

[7] *Tables of the Cumulative Binomial Probabilities*, U. S. Army Ordnance Corps, Ordnance Corps Pamphlet ORDP 20-1.

A SMOOTH INVERTIBILITY THEOREM[1]

By John W. Tukey

Princeton University

1. Introduction. In connection with discussions of fiducial inference (e.g., see 3), it is often desirable to consider the invertibility of certain mappings We shall say that a mapping is smoothly invertible (of class α) if (condition (1) of [3] is irrelevant here):

(2) the mapping is 1-1 and hence has a single-valued inverse,

(3) this inverse is a continuous function (and has continuous derivatives of all orders up to α).

All too often, as has been emphasized to the author by L. J. Savage, the question of invertibility has been "answered" by showing that a Jacobian is of constant sign. It is, of course, well known that this does not suffice to give uniqueness in the large.

Explicit conditions sufficient for uniqueness in the large do not seem to be given frequently in the literature. The present note records an explicit theorem in a form which seems likely to be of service in such conditions.

2. A smooth invertibility theorem. We now state the smooth invertibility theorem as follows:

Any α times continuously differentiable mapping from an arcwise connected open domain (in n dimensions) *to a simply connected range, whose Jacobian determinant is continuous and of one sign throughout the domain, and whose inverse carries compact sets into compact sets is smoothly invertible of order α.*

In our application it is convenient to use the

Observation. *If the open domain and the simply connected range are both the whole plane (or the whole of any Euclidian space) then the inverse will carry compact sets into compact sets provided that it carries bounded sets into bounded sets.*

The proof of this observation follows immediately from the remarks that (i) the inverse image of closed sets by a continuous mapping are always closed, (ii) in the whole plane the compact sets are just those which are closed and bounded.

3. Proof. The proof of the theorem rests on a classical result about local inversion (which makes no use of arcwise connectedness, simple connectedness, or the hypothesis about the inverse taking compact sets into compact sets) and a purely topological result relating local uniqueness of inverses to global uniqueness (which makes no explicit use of the differentiability conditions)

We say that N is a local inverse neighborhood of x if

(1) $f(N)$ is a neighborhood of $y = f(x)$,

Received March 4, 1957.

[1] Prepared in connection with research sponsored by the Office of Naval Research

(2) there is a choice $g_N(y')$ defined for y' in $f(N)$ such that $f(g_n(y')) = y'$ for any y' in $f(N)$.

If there is but one inverse image x' in N for each y' in $f(N)$, N is a unique local inverse neighborhood. If every x has a unique local inverse neighborhood, we say that $y = f(x)$ has unique local inverses.

The classical result (given for example in [1] on pp. 257–258 of Vol. 2) asserts that under our differentiability and Jacobian hypothesis, $y = f(x)$ has unique local inverses, and that, if we restrict the neighborhoods sufficiently, these inverses are α times continuously differentiable.

All that is lacking is the knowledge that these local inverses are unique, not only when we must go back to a neighborhood N of a particular solution, x of $y = f(x)$, but when we are free to go back to any x. This will follow from a special case of the covering homotopy theorem, which will be derived from a more usual form in the next section, namely: *If $y = f(x)$ is continuous and has unique local inverses, if X is a compact metric space, if x_t is a continuous image of the unit interval $0 \leqq t \leqq 1$ in X, if $y_{t,s}$ is a continuous image of the unit square $0 < t, s < 1$ in $f(X)$, and if $y_{t,1} = f(x_t)$, then it is possible to define a continuous image $x_{t,s}$ of the unit square in X so that*

$$x_{t,1} = x_t , \tag{1}$$

$$f(x_{t,s}) = y_{t,s} \tag{2}$$

(and indeed this can be done in at most one way.)

Using this result we can complete the proof of the smooth invertibility theorem as follows: Let x_0 and x_1 be any two solutions (possibly coincident) of $y_0 = f(x)$. Since X is arcwise connected, we may join x_0 to x_1 by an arc x_t for $0 \leqq t \leqq 1$. Let $y_{t,1} = f(x_t)$. Since $y_{0,1} = y_{1,1} = y_0$ the image of this arc is a closed curve (in $f(X)$). Since $f(X)$ is simply connected, this curve can be shrunk to the point y_0, keeping its ends at y_0; that is to say, we can define $y_{t,s}$ for $0 \leqq t, s < 1$ as a continuous extension of $y_{t,1}$ with $y_{t,0} \equiv y_{0,s} \equiv y_{1,s} \equiv y_0$. Let H be the set of all y of the form $y_{t,s}$ for $0 \leqq t, s < 1$. As the continuous image of a compact space this will be compact and will hence be closed in $f(X)$. Let G be the set of x for which $y = f(x)$ lies in H. Because of our hypothesis, G will also be compact. Surely G contains x_t for $0 \leqq t \leqq 1$.

Now apply the topological result to G. We have then a continuous image $x_{t,s}$ of $0 \leqq t, s \leqq 1$ which satisfies $x_{0,1} = x_0$, $x_{1,1} = x_1$, and $f(x_{t,0}) \equiv f(x_{0,s}) \equiv f(x_{1,s}) = y_0$.

The images of the three sides $s = 0$, $t = 0$, and $t = 1$ of the unit square thus provide an arc leading from x_0 to x_1 every point of which maps into y_0. The local uniqueness of inverses now ensures that this arc is a constant mapping, and hence that, in particular $x_0 = x_1$. Thus the solution of $y_0 = f(x)$ is shown to be unique, and the proof of the smooth invertibility theorem is concluded.

There is some interest in the need for all the topological hypotheses of this theorem, so examples are given in Section 5 to show that no one of these three hypotheses can be removed without the conclusion failing.

4. Proof of topological result. The textbook of Seifert and Threlfall [2] gives, as Satz I and Satz II (pp. 186–188), theorems from which we can immediately derive a close analog of the result we used. The differences will be as follows: (i) the result will be restricted to a topological complex (which would suffice, since we wish to apply it only to n-dimensional open domains), (ii) the condition that the inverse image of compact sets is compact would be absent, (iii) the condition of unique local inverses would be strengthened to local homeomorphism. We need then only show that our extra condition implies local homeomorphism.

Let y be a point of $f(X)$. Let $x_1, x_2, \cdots, x_k$ be its inverse images. Let $N_1, N_2, \cdots, N_k$ be corresponding unique local inverse neighborhoods. Let K be the part of X not in any of these N_i. Consider L, the complement of the closure of $f(K)$. If y is in L, then the intersection of L with all $f(N_i)$ is a neighborhood of y each of whose points has exactly k inverse images, one each in $N_1, N_2, \cdots, N_k$, and since these local inverses are continuous, we have the desired local homeomorphism.

If y is not in L, then it is in $\bar{f}(K)$, the closure of $f(K)$, and there is a sequence y_i in $f(K)$ which converges to y. Let $Y = y, y_1, y_2, \cdots$. Then Y is compact.

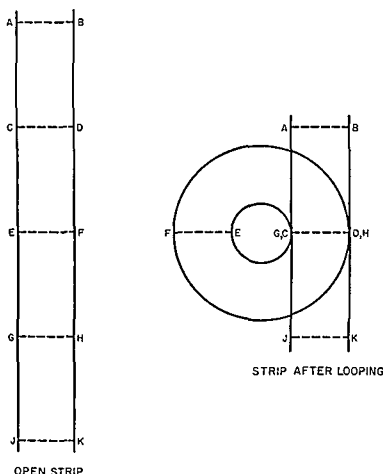


Fig. 1. Second step in the mapping

Let x_j in K satisfy $f(x_j) = y_j$. Then x_j is in the inverse image of Y which is compact. Hence we can extract a subsequence x_{j_n} which converges to some x_0. Clearly $f(x_0) = \lim f(x_{j_n}) = \lim y_{j_n} = y$. Hence x_0 is in some N_j, and hence not in K, which is a contradiction. The desired result is thus proved.

5. Remarks and counter examples. In this section we show that no one of the three topological hypotheses can be omitted. In each case we show that uniqueness of the inverse is immediately lost. For the first two hypotheses the example can be very simple and one-dimensional. For the third a two-dimensional, not-too-simple example is provided.

If we drop the connectedness requirement for X, then we may take X as two non-intersecting straight lines and $f(x)$ as a rigid application of each on a third. Uniqueness is lost. (Consideration shows, indeed, that ϵ-connectedness for all $\epsilon > 0$ would suffice.)

If we drop the simple connectedness of $f(X)$, then we may take X as a circle, $f(X)$ as a circle of half that diameter, and the mapping $f(X)$, as the winding of an inextensible loop of string around the smaller circle. See Fig. 1. (We cannot use the whole infinite line without having the inverse image of a compact set become non-compact.)

In the third example we drop the requirement on compactness of inverse images. Here the example is a mapping of the plane into part of itself which is most easily described qualitatively and geometrically. We begin by deforming the plane into a long open strip, which can clearly be done with a positive, non-zero Jacobian. We now consider a transformation of the strip into a simply looped strip in which uniqueness of inverse has been lost but simple connectedness of range has not been achieved. Graphically, corresponding points appear as in Fig. 1. The failure of simple connectivity is due to the small circular disk bounded by the image of the arc CEG. If we loop the strip again in the same way, we can use the new loop to cover this hole, meanwhile placing the new hole on the old loop. The resulting transformation will have a simply connected range, but the inverses will not be unique. This is the desired third example.

REFERENCES

[1] A. Ostrowski, *Vorlesungen uber Diffential und Integralrechnung*, Basel, Birkhauser, 1951.

[2] H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Leipzig, Taubner, 1938, and New York, Chelsea, 1947.

[3] John W. Tukey, "Some examples with fiducial relevance," *Ann. Math. Stat.*, Vol. 28, (1957), pp. 687–695.

ON THE INVERSION OF THE SAMPLE COVARIANCE MATRIX IN A STATIONARY AUTOREGRESSIVE PROCESS[1]

By M. M. Siddiqui[2]

University of North Carolina

Let $x_1, \cdots, x_N$ be the observations on a variate at times $t = 1, \cdots, N$. It is assumed that the underlying model is an autoregressive scheme of order k

(1) $$a_0 x_t + a_1 x_{t-1} + \cdots + a_k x_{t-k} = z_t,$$

where z's are independent $N(0, 1)$ variates and the roots of the equation

$$\sum_{j=0}^{k} a_j y^j = 0$$

lie inside the unit circle $|y| = 1$ in the complex plane. The variate z_t is, then, independent of $x_{t-1}, x_{t-2}, \cdots$ ([2], p. 38). It is further assumed that the process is stationary so that $E\, x_t$, $E\, x_t x_{t+j}$, $j = 0, 1, 2, \cdots$ are independent of t. Writing σ_x^2 for the variance of any x, we observe that since $E\, x_t = 0$, $\sigma_x^2 = E\, x_t^2$. We define autocorrelation between x_t and x_s by

(2) $$\gamma_{|t-s|} = E\, x_t x_s / \sigma_x^2$$

so that γ_t satisfies Eq. (1) with z_t replaced by zero and $\gamma_{-t} = \gamma_t$.

Let X_j stand for the column vector of the first j observations and X_j' for its transpose, i.e.,

(3) $$X_j' = (x_1, \cdots, x_j), \qquad j = 1, 2, \cdots, N.$$

Also write A_j for the covariance matrix of the vector X_j, i.e.,

(4) $$A_j = \sigma_x^2 \begin{bmatrix} 1 & \gamma_1 & \gamma_2 & \cdots & \gamma_{j-1} \\ \gamma_1 & 1 & \gamma_1 & \cdots & \gamma_{j-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{j-1} & \gamma_{j-2} & \gamma_{j-3} & \cdots & 1 \end{bmatrix},$$

for $j = 1, 2, \cdots, N$. We note here that the matrix A_j is persymmetric, i.e., symmetric about both the diagonals. This property will be used to obtain A_N^{-1}

The distribution of X_N is given by

(5) $$dF(X_N) = (2\pi)^{-N/2} |A_N|^{-1/2} \exp\left[-\tfrac{1}{2}(X_N' A_N^{-1} X_N)\right] dX_N.$$

J. Wise [1] has given a method of finding A_N^{-1} using the spectral density function. We propose here another method of obtaining A_N^{-1} based on the symmetric property of A_N.

Received February 2, 1957; revised November 13, 1957.

[1] Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] At present with the Boulder Laboratories, National Bureau of Standards

The distribution of $x_1, \cdots, x_k, z_{k+1}, \cdots, z_N$ is given by

$$dF(x_1, \cdots, x_k, z_{k+1}, \cdots, z_N) = (2\pi)^{-N/2} | A_k |^{-1/2} \exp\left[-\frac{1}{2}\left\{ X_k' A_k^{-1} X_k + \sum_{t=k+1}^{N} z_t^2 \right\}\right] dX_k \, dz_{k+1} \cdots dz$$

We shall assume here that $N > 2k$. Considering (1) as a transformation fr z_t to x_t for $t = k + 1, \cdots, N$, we obtain the distribution of X_N as

$$dF(X_N) = (2\pi)^{-N/2} a_0^{N-k} | A_k |^{-1/2} \cdot \exp\left[-\frac{1}{2}\left\{ X_k' A_k^{-1} X_k + \sum_{t=k+1}^{N} \left(\sum_{i=0}^{k} a_i x_{t-i} \right)^2 \right\}\right] dX \tag{6}$$

Comparing (5) and (6) we have

$$a_0^{2N} | A_N | = a_0^{2k} | A_k | \tag{7}$$

and

$$X_N' A_N^{-1} X_N = X_k' A_k^{-1} X_k + \sum_{t=k+1}^{N} \left(\sum_{i=0}^{k} a_i x_{t-i} \right)^2 . \tag{8}$$

Let C_N be the $N \times N$ matrix which has A_k^{-1} in the upper right-hand corn and zeroes elsewhere, i.e.,

$$C_N = \begin{bmatrix} A_k^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \tag{9}$$

and B_N be the matrix of the quadratic form in the second term on the right Eq. (8), i.e.,

$$\sum_{t=k+1}^{N} \left(\sum_{i=0}^{k} a_i x_{t-i} \right)^2 = X_N' B_N X_N , \tag{10}$$

so that we have

$$A_N^{-1} = B_N + C_N . \tag{11}$$

Denoting by a_{ij}, b_{ij}, and c_{ij}, $i, j = 1, 2, \cdots, N$, the elements in the i row and jth column of the matrices A_N^{-1}, B_N, and C_N respectively, we have

$$a_{ij} = b_{ij} + c_{ij} . \tag{12}$$

But $c_{ij} = 0$ if either i or $j > k$. Hence

$$a_{ij} = b_{ij} \text{ if either } i \text{ or } j > k. \tag{13}$$

Now B_N is completely known. In fact, assuming $j \geqq i$,

$$b_{ji} = b_{ij} = \begin{cases} \sum_{t=k+1}^{k+i} a_{t-i} a_{t-j} & \text{for } j \leqq k, \\ 0 & \text{for } i + k < j \leqq N, \quad i \leqq N - k, \\ \sum_{t=j}^{k+i} a_{t-i} a_{t-j} & \text{for } k + 1 \leqq j \leqq i + k, \quad i \leqq N - \\ \sum^{N} a_{t-i} a_{t-j} & \text{for } i \geq N - k + 1 \end{cases} \tag{14}$$

Thus all the a_{ij}, except those for which both i and j are less than or equal to k, are known. Now, since A_N is persymmetric, so is A_N^{-1}. Therefore

$$a_{ji} = a_{ij} = a_{N-i+1,N-j+1}, \qquad i, j = 1, 2, \cdots, N. \tag{15}$$

Using (13) and remembering that $N > 2k$, we have

$$a_{ij} = a_{ji} = b_{N-i+1,N-j+1} \qquad \text{for } i, j = 1, 2, \cdots, k. \tag{16}$$

Thus A_N^{-1} is completely determined. We now use relations (12) to obtain all the elements of A_k^{-1}. Once A_k^{-1} is known, we can find A_N^{-1} for any $N \geqq k$ using (6).

If k is less than 5, we can directly compute A_k^{-1} and then use Eq. (6) to obtain A_N^{-1}.

Illustration. Let $k = 2$. The distribution of X_N is

$$dF(X_N) = (2\pi)^{-N/2} a_0^{N-2} |A_2|^{-1/2} \cdot \exp\left[-\frac{1}{2}\left\{X_2' A_2^{-1} X_2 + \sum_{t=3}^{N} (a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2})^2\right\}\right] dX_N,$$

so that

$$B_N = \begin{bmatrix} a_2^2 & a_1 a_2 & a_0 a_2 & 0 & \cdots & 0 & 0 \\ a_1 a_2 & a_1^2 + a_2^2 & a_0 a_1 + a_1 a_2 & a_0 a_2 & \cdots & 0 & 0 \\ a_0 a_2 & a_0 a_1 + a_1 a_2 & a_0^2 + a_1^2 + a_2^2 & a_0 a_1 + a_1 a_2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & a_0^2 + a_1^2 & a_0 a_1 \\ 0 & 0 & 0 & 0 & \cdots & a_0 a_1 & a_0^2 \end{bmatrix}.$$

Hence

$$A_N^{-1} = \begin{bmatrix} a_0^2 & a_0 a_1 & a_0 a_2 & 0 & \cdots & 0 & 0 \\ a_0 a_1 & a_0^2 + a_1^2 & a_0 a_1 + a_1 a_2 & a_0 a_2 & \cdots & 0 & 0 \\ a_0 a_2 & a_0 a_1 + a_1 a_2 & a_0^2 + a_1^2 + a_2^2 & a_0 a_1 + a_1 a_2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & a_0^2 + a_1^2 & a_0 a_1 \\ 0 & 0 & 0 & 0 & \cdots & a_0 a_1 & a_0^2 \end{bmatrix},$$

$$A_2^{-1} = \begin{bmatrix} a_0^2 - a_2^2 & a_0 a_1 - a_1 a_2 \\ a_0 a_1 - a_1 a_2 & a_0^2 - a_2^2 \end{bmatrix},$$

$$|A_2^{-1}| = (a_0^2 - a_2^2)^2 - (a_0 - a_2)^2 a_1^2,$$

and

$$a_0^{2N} \mid A_N \mid = a_0^4[(a_0^2 - a_2^2)^2 - (a_0 - a_2)^2 a_1^2]^{-1}.$$

It may be mentioned here that Ulf Gernander and Murray Rosenblatt ([2], pp. 238–239) have considered asymptotic properties of A_N^{-1} as N tends to infinity. They, however, do not attempt to determine the k^2 elements standing in the first k rows and the first k columns of A_N^{-1}, although they suggest a method of orthogonalization of the vector X_N.

REFERENCES

[1] J. Wise, "The autocorrelation function and the spectral density function," *Biometrika*, Vol. 42 (1955), pp. 151–159.
[2] Ulf Gernander and Murray Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1957.

A PROBLEM OF BERKSON, AND MINIMUM VARIANCE ORDERLY ESTIMATORS[1]

By John W. Tukey

Princeton University

1. Summary. The distinction between efficiency in the asymptotic sense originally introduced by Fisher ([2], 1925, p. 703), and the finite sample sense sometimes used by others has been recently stressed by various writers (e.g., Berkson [1]). The technique of proof used below was originally developed to provide a simple example where the maximum likelihood estimate of location, though asymptotically efficient, was not of minimum variance for any finite sample size whatever. The (symmetrical) double exponental distribution with known scale, where the sample median is the maximum likelihood estimator of location, could easily be shown to be such an example. (While this result is useful in deflating unwarranted views about minimum variance properties of maximum likelihood estimates, Fisher's ([2], p. 716) results about intrinsic accuracy in the same situation are of more basic interest.)

On examination, however, the technique used to provide this rather isolated and special result was found capable of showing, for a class of distributions with suitable monotony properties (in particular all distributions for which $f'(y)/f(y)$ is monotone decreasing, and all normal, exponential, gamma and beta distributions), that the covariances of the order statistics in a sample of any chosen size are monotone in either index separately.

Received March 22, 1957.

[1] Prepared in connection with research sponsored by the Office of Naval Research. Based, in part, on Memorandum Report 11 of the Statistical Research Group, Princeton University, which was issued 25 October 1948.

2. Complete regression. We shall say the distribution of z given y *shows complete negative regression on* y if the cumulative distribution $F(z \mid y)$ satisfies

$$F(z \mid y'') \leqq F(z \mid y') \qquad \text{for } y'' \leqq y',$$

provided the equality does not always hold. We define complete positive regression analogously. We notice that:

(A_1) *If the distribution of z given y shows complete negative regression y, and z_1 is an order statistic from a random sample of z's, then the distribution of z_1 given y shows complete negative regression on y.*

(A_2) *If the distribution of w given z shows complete negative regression on z, and the distribution of z given y shows complete positive regression on x, and the distribution of w given z is unaffected by giving y, then the distribution of w given y shows complete negative regression on y.*

(A_3) *If the distribution of z given y shows complete negative regression on y then* cov $\{z, y\} < 0$.

The first result follows from the beta-function formula for an order statistic cumulative,

$$G_{k|n}(w) = k\binom{n}{k}\int_0^{F(w)} t^{k-1}(1-t)^{n-k}\,dt$$

which follows from the interpretation of $G_{k|n}(w)$ as the probability of k or more out of n falling in an interval of probability $F(w)$, and which shows that the cumulative of $z_{k|n}$ is a monotone function of the cumulative of z. The second follows easily by introducing the monotone *representing function* [3] $z_y(u)$ corresponding to $F(z \mid y)$ such that if u is uniformly distributed on $[0, 1]$, then $z_y(u)$ is distributed according to $F(z \mid y)$. The hypothesis of complete positive regression is equivalent to $z_{y'}(u) \geqq z_{y''}(u)$ for $y' \geqq y''$, and we have

$$H(w \mid y') = \int G(w \mid z_{y'}(u))\,du \leqq \int G(wz_{y''}(u))\,du = H(w \mid y'') \qquad \text{for } y' \leqq y''$$

which we desired to show. The third result follows from the fact that

$$\text{ave}\,\{z \mid y'\} \geqq \text{ave}\,\{z \mid y''\}$$

which follows directly from the inequality for the representing functions

3. Subexponential distributions. We shall say that a cumulative distribution is *subexponential to the right* if

$$\frac{F(z+h) - F(h)}{1 - F(h)} = 1 - \frac{1 - F(z+h)}{1 - F(h)}$$

is monotonically decreasing for fixed $z > 0$ as h increases. We notice that this is equivalent to stating that, *referred to the point of truncation*, the distribution of z after truncation on the left shows complete negative regression on the

point of truncation. We define subexponential on the left, or in both directions, analogously. We are now prepared to demonstrate:

(B$_1$) *If* $F(q) = F(y - 0)$ *is subexponential on the right, and* $y_1 \leqq y_2 \leqq \cdots \leqq y_n$ *are an ordered sample of* y's (*and* $q_1 \leqq q_2 \leqq \cdots \leqq q_n$ *are an ordered sample of* q's), *then for any* $j < k$ *we have*

$$\text{cov}\,\{y_k - y_j, y_j\} = \text{cov}\,\{q_k - q_j, q_j\} < 0.$$

(The analogs for "on the left" or "in both directions" clearly follow by symmetry.) The proof rests on Wald's principle ([4], p. 536) according to which the distribution of q_k given q_j is that of the $(k - j - 1)$st order statistic from a sample of $n - j$ from the result of truncating the original distribution at q_j. The distribution of $q_k - q_j$ for q_j fixed is that of a similar order statistic from the truncated distribution referred to its point of truncation as origin—and as remarked above this latter distribution shows complete negative regression on q_j. By (A$_1$) the same is true of any distribution of an order statistic, and hence for the distribution of $q_k - q_j$. The negativity of the covariance follows from (A$_3$).

This result (and its analogs) can easily be extended to

(B$_2$) *Under the hypotheses of* (B$_1$), *if* $h \leqq j < k$, *then*

$$\text{cov}\,\{y_k - y_j, y_h\} = \text{cov}\,\{q_k - q_j, q_h\} < 0.$$

For, since the distribution of q_k given q_j is not affected by giving q_h, and the distribution of q_j given q_h shows complete positive regression on q_h, we may apply (A$_2$) to complete the proof.

As a corollary we have the curiously simple results:

(B$_3$) *If the distribution of* q *is subexponential in both directions, then the covariance of any two order statistics is less than the variance of either.*

(B$_4$) *If the distribution of* q *is subexponential in both directions, the covariance between order statistics* q_j, q_k *is monotone in* j *and* k *separately, decreasing as* j *and* k *separate from one another.*

The interest of these results is enhanced when we observe normal, exponential, gamma and beta distributions, pristine or truncated, are all subexponential in both directions.

4. Monotone location-scores. By definition, a distribution $F(q)$ is subexponential to the right if

$$G(y \mid h) = \frac{F(y + h) - F(h)}{1 - F(h)} = 1 - \frac{1 - F(y + h)}{1 - F(h)}$$

is monotone decreasing as h increases for every y. This is equivalent to

$$\log\,\{1 - F(y + h)\} - \log\,\{1 - F(h)\}$$

being monotone increasing, or, granting differentiability, to

$$\frac{f(h)}{1 - F(h)} - \frac{f(y + h)}{1 - F(y + h)} > 0,$$

where $y > 0$, and hence to

$$-\frac{f(u)}{1 - F(u)} = \frac{d}{du} \log (1 - F(u)) = \frac{\int_u^\infty f'(u)\, du}{\int_u^\infty f(u)\, du}$$

being monotone decreasing. This will follow from the monotone decreasing character of $(\log f(u))' = f'(u)/f(u)$ since $f(u) \geqq 0$. It is thus sufficient, but not necessary, for subexponentiality on the right that $f'(u)/f(u)$, which is the location score associated with the specification consisting of all translations of $F(u)$, be monotone decreasing.

The class of distributions with monotone scores for location is immediately seen to be closed under formal multiplication of densities, so that if

(i) $F(u)$, $G(u)$ have monotone scores for location,

(ii) $F(u) = \int f(u)\, du,\ G(u) = \int g(u)\, du,$

(iii) $H(u) = (\text{constant})\, f(u)g(u)\, du,$

then $H(u)$ also has a monotone score for location. The class is also closed under truncation at one or both ends. It is immediately seen to include all distributions whose shapes are single exponential, double exponential (balanced or not), normal, (incomplete) gamma, (incomplete) beta, and their formal products and truncations. (It does *not* include distributions of Cauchy shape.)

Since the large-sample optimum weight to be assigned to an order statistic is the negative of the derivative of the score for location at the typical point of distribution, it seems both peculiarly appropriate and highly reasonable that the minimum variance orderly estimator of location will actually have all its coefficients positive for any distribution with monotone score for location.

5. Orderly estimates. We now turn to a location specification $F(y \mid \theta) = F(y - \theta)$ and to orderly estimates of θ, by which we mean linear combinations of order statistics of total weight 1, i.e.,

$$\tilde{y} = \sum w_i y_i + c, \qquad \sum w_i = 1.$$

(Notice that the variance and bias of $\tilde{y}$ as an estimator of θ are exactly the variance and average value of $\tilde{q}$, where $\tilde{q} = \sum w_i q_i$ are order statistics in a sample from $F(q)$.) We begin with a general result, applicable to *any* convex (i.e., closed under $\alpha t' + (1 - \alpha)t''$ for $0 < \alpha < 1$) class of estimates of u which contains all order statistics (and thus surely contains all orderly estimates).

(C) *If t is the minimum variance estimate in any convex class containing the order statistics, and y_k is any order statistic (or any linear combination of order statistics of total weight one)*

$$\operatorname{cov} \{y_k - t, t\} = 0.$$

This follows easily by considering the variance of $t + \lambda(y_k - t)$, where, in par-

ticular, if cov $\{y_k - t, t\} < 0$, a value of $\lambda > 0$ will provide lesser variance than for $\lambda = 0$.

From the result it is easy to show that:

(D) *If $F(q)$ is subexponential to the right, then no single order statistic, except possibly the righthandmost, is of minimum variance among orderly estimates of location.*

(Again, the analogs with "to the left ... the lefthandmost" or "in both directions ... statistic," follow by symmetry.) For if y_j were of minimum variance, and y_n the righthandmost, then by

$$(B_1) \text{ cov } (y_n - y_j, y_j) = \text{cov } (q_n - q_j, q_j) < 0,$$

and by (C) y_j is not of minimum variance. It is reasonable to anticipate that, actually, all coefficients must be positive (particularly for distributions with monotone scores), but the elementary methods used here do not seem to show this easily.

REFERENCES

[1] J. Berkson, "Question 21," *Amer. Statistician*, Vol. 3 (Oct., 1949), p. 12.
[2] R. A. Fisher, "Theory of statistical estimation," *Proc. Cambridge Philos. Soc.*, Vol. 22 (1923–25), pp. 700–725.
[3] C. Hastings, F. Mosteller, J. W. Tukey and C. P. Winsor, "Low moments for small samples: a comparative study of order statistics," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 413–426.
[4] J. W. Tukey, "Non-parametric estimation, II. Statistical equivalent blocks and tolerance regions ... the continuous case," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 529–539.

AN ELEMENTARY THEOREM CONCERNING STATIONARY ERGODIC PROCESSES

By Leo Breiman

University of California, Berkeley

1. Introduction. The purpose of this note is to state and prove a theorem concerning strictly stationary, ergodic processes and to give some of its applications. Although the theorem itself is a simple consequence of the ergodic theorem, its applications include a proof of the consistency of the maximum likelihood estimates for stationary distributions and an extension of the zero-one law for symmetric sets given by Hewitt and Savage [1].

Theorem 1. *Let $\cdots x_{-1}, x_0, x_1, \cdots$ be a strictly stationary process such that every set invariant under shifts has measure zero or one. Let $\{\phi_n\}$ be a sequence of real-valued functions, ϕ_n being a measurable function of $n + 1$ variables. Then if the sequence $\phi_n(x_0, \cdots, x_n)$ and the sequence $\phi_n(x_{-n}, \cdots, x_0)$ both converge in probability, their limits are almost surely constant and equal.*

Received June 25, 1957; revised October 25, 1957.

PROOF. We assume that $\phi_n(x_{-n}, \cdots, x_0) \xrightarrow{P} \phi^-$, $\phi_n(x_0, \cdots, x_n) \xrightarrow{P} \phi^+$. Let f be any differentiable function such that $|f| \leqq 1$, $|f'| \leqq 1$, and let

$$f_n^- = f(\phi_n(x_{-n}, \cdots, x_0)), \qquad f_n^+ = f(\phi_n(x_0, \cdots, x_n)).$$

Since $|f_n^- - f(\phi^-)| \leqq |\phi_n(x_{-n}, \cdots, x_0) - \phi^-|$, we have that $f_n^- \xrightarrow{P} f(\phi^-)$. From the uniform boundedness of the f_n^-, it follows that $\int |f_n^- - f(\phi^-)| \to 0$, and similarly $\int |f_n^+ - f(\phi^+)| \to 0$. We denote by T the shift operation, so that

$$T^n\phi_n(x_n, \cdots, x_0) = \phi_n(x_0, \cdots, x_n).$$

By the ergodic theorem,

$$\lim_n \int \left| Ef(\phi^-) - \frac{1}{n}\sum_{k=1}^n T^k f(\phi^-) \right| = 0.$$

Hence,

$$\int |Ef(\phi^-) - f(\phi^+)| \leqq \limsup_n \frac{1}{n}\sum_1^n \int |T^k f(\phi^-) - T^k f_k^-| + \limsup_n \frac{1}{n}\sum_1^n \int |f_k^+ - f(\phi^+)|.$$

Since T is measure preserving, the terms on the right may be reduced to

$$\limsup_n \frac{1}{n}\sum_1^n \int |f(\phi^-) - f_k^-| + \limsup_n \frac{1}{n}\sum_1^n \int |f_k^+ - f(\phi^+)| = 0.$$

We conclude that $f(\phi^+)$ is almost surely constant, which proves the theorem.

2. Applications. We use the above theorem first to prove a result concerning maximum likelihood ratios, which is a special case of a theorem due to C. Kraft [2].

THEOREM 2. *Let the $\cdots x_{-1}, x_0, x_1, \cdots$ process be distributed according to the stationary ergodic measure P with density functions $p_n(\cdot, \cdots, \cdot)$ and let Q be any other stationary measure with density functions $q_n(\cdot, \cdots, \cdot)$ such that P is not absolutely continuous with respect to Q. Then almost surely (a.s.),*

$$\lim_n \frac{q_n(x_0, \cdots, x_n)}{p_n(x_0, \cdots, x_n)} = 0.$$

PROOF. Let $\phi_n = q_n(\cdot, \cdots, \cdot)/p_n(\cdot, \cdots, \cdot)$; it is well known ([3], pp. 93, 348) that the sequence $-\phi_n(x_0, \cdots, x_n)$ forms a semi-martingale with respect to the fields B_{n-1}^+ generated by $x_0, \cdots, x_{n-1}$. Similarly, the sequence

$$-\phi_n(x_{-n}, \cdots, x_0)$$

forms a semi-martingale with respect to the fields B_{n-1}^- generated by $x_{-n+1}, \cdots, x_0$. Since in both cases the first absolute moments are bounded by one, both sequences converge a.s. From our main theorem we conclude that there is some constant α such that

$$\lim_n \frac{q_n(x_0, \cdots, x_n)}{p_n(x_0, \cdots, x_n)} = \alpha \quad \text{(a.s.)}.$$

Now for any finite dimensional cylinder set I, we have for all n sufficiently large,

$$Q(I) \geqq \int_I \phi_n(x_0, \cdots, x_n)\, dP.$$

Equality would obtain, except that p_n may vanish on some set of positive measure where q_n does not. Using the Fatou lemma,

$$Q(I) \geqq \liminf \int_I \phi_n(x_0, \cdots, x_n)\, dP \geqq \int_I \liminf \phi_n(x_0, \cdots, x_n)\, dP = \alpha P(I).$$

Since the above is true for any cylinder set, it is also true for any finite disjoint union of cylinder sets, and thus is true in general, contradicting our hypothesis unless $\alpha = 0$.

Another application of Theorem 1 results in an extremal property of random variables having symmetric distributions. While we do not know of an explicit statement of this theorem, it can also be proven using de Finetti's representation theorem (see, for example [1] and [4], p. 364), without too much difficulty.

THEOREM 3. *Let $x_0, x_1, \cdots$ be a sequence of random variables whose finite dimensional distribution functions are invariant under permutations of the arguments and such that every "tail" event has measure zero or one. Then the sequence is equivalent to a sequence of independent random variables.*

PROOF. From the symmetry of the $x_0, x_1, \cdots$ sequence follows its stationarity. By the usual procedure we extend the measure to the double-ended sequence $\cdots x_{-1}, x_0, x_1, \cdots$, noticing that the symmetry is preserved under this extension. We also verify that the zero-one hypothesis implies the process is ergodic so that we may apply Theorem 1. We define $\phi_n(x_0, x_1, \cdots, x_n) = p(x_{-1} \leqq a \mid x_0, \cdots, x_n)$. Then, by symmetry and stationarity

$$\phi_n(x_{-n}, \cdots, x_0) = p(x_1 \leqq a \mid x_0, \cdots, x_{-n}).$$

By the martingale convergence theorem, both

$$\phi_n(x_0, \cdots, x_n),\ \phi_n(x_{-n}, \cdots, x_0)$$

converge a.s. and we conclude that both

$$p(x_{-1} \leqq a \mid x_0, x_1, \cdots),\ p(x_1 \leqq a \mid x_0, x_{-1}, \cdots)$$

are a.s. constant, which proves the theorem.

A more specialized consequence of Theorem 1 runs as follows.

THEOREM 4. *Let $x_0, x_1, \cdots$ be a sequence of identically distributed, independent random variables, and $\{\phi_n\}$ a sequence of real-valued functions, ϕ_n a measurable function of $n + 1$ variables. Then if both $\phi_n(x_0, \cdots, x_n)$ and*

$$\phi_n(x_n, \cdots, x_0)$$

converge in probability, their limits are a.s. constant.

PROOF. By the usual procedure we extend the measure on $x_0, x_1, \cdots$ to a measure on the two-sided process $\cdots, x_{-1}, x_0, x_1, \cdots$. The set

$$[|\phi_m(x_{-m}, \cdots, x_0) - \phi_n(x_{-n}, \cdots, x_0)| > \epsilon]$$

has the same measure as the set $[|\phi_m(x_m, \cdots, x_0) - \phi_n(x_n, \cdots, x_0)| > \epsilon]$. Hence $\phi_n(x_{-n}, \cdots, x_0)$ converges in probability and Theorem 1 applies

The above theorem is an extension of the Hewitt-Savage zero-one law for symmetric sets, as the following theorem makes clear.

THEOREM 5. *Let $x_0, x_1, \cdots$ be a sequence of identically distributed, independent random variables and f any integrable function on the process such that f is invariant under finite permutations of the coordinates. Then f is a.s. constant.*

PROOF. Let $\phi_n(x_0, \cdots, x_n) = E(f \mid x_0, \cdots, x_n)$. Then $\phi_n(x_n, \cdots, x_0) = \phi_n(x_0, \cdots, x_n)$ by the symmetry of f and the $\phi_n(x_0, \cdots, x_n)$ sequence forms a martingale which converges a.s. to f. The conclusion follows from Theorem 3.

REFERENCES

[1] E. HEWITT AND L. J. SAVAGE, "Symmetric measures on Cartesian products," *Trans. Amer. Math Soc*, Vol. 80 (1955), pp. 470–501.
[2] C. KRAFT, "Some conditions for consistency and uniform consistency of statistical procedures," *Univ. of California Publ. Stat*, Vol. 2, No. 6 (1955), pp. 125–141.
[3] J. L DOOB, *Stochastic Processes*, John Wiley and Sons, New York, 1953.
[4] M. LOÈVE, *Probability Theory*, D. Van Nostrand, New York, 1955.

A TEST OF FIT FOR MULTIVARIATE DISTRIBUTIONS[1]

BY LIONEL WEISS

Cornell University

1. Summary and introduction. Suppose X is a chance variable taking values in k-dimensional Euclidean space. That is, $X = (Y_1, \cdots, Y_k)$, where Y_i is a univariate chance variable. The joint distribution of $(Y_1, \cdots, Y_k)$ has density $f(y_1, \cdots, y_k)$, say.

We shall call a function $h(y_1, \cdots, y_k)$ "piecewise continuous" if it is everywhere bounded, and k-dimensional Euclidean space can be broken into a finite number of Borel-measurable subregions, such that in the interior of each subregion $h(y_1, \cdots, y_k)$ is continuous, and the set of all boundary points of all subregions has Lebesgue measure zero.

We assume that $f(y_1, \cdots, y_k)$ is piecewise continuous. Let $h(y_1, \cdots, y_k)$ be some given nonnegative piecewise continuous function, and let $X_1, \cdots, X_n$ be independent chance variables, each with the density $f(y_1, \cdots, y_k)$. Choose a nonnegative number t, and for each i, construct a k-dimensional sphere with center at $X_i = (Y_{i1}, \cdots, Y_{ik})$ and of k-dimensional volume

$$\frac{th(Y_{i1}, \cdots, Y_{ik})}{n}.$$

Such a sphere will be called "of type s" if it contains exactly s of the $(n - 1)$

Received August 8, 1957.
[1] Research sponsored by the Office of Naval Research.

points $X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n$. Let $R_n(t; s)$ denote the proportion of the n spheres which are of type s.

For typographical simplicity, we denote the vector $(y_1, \cdots, y_k)$ by y. Let $S(t; s)$ denote the multiple integral

$$(t^s/s!) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h^s(y) f^{s+1}(y) \exp\{-th(y)f(y)\}\, dy_1 \cdots dy_k.$$

It is shown that $R_n(t; s)$ converges stochastically to $S(t; s)$ as n increases. This result is then used to construct a test of the hypothesis that the unknown density $f(y)$ is equal to a given density $g(y)$.

2. Proof of the convergence of $R_n(t; s)$. We define the chance variable Z_i to be equal to one if the sphere centered at X_i is of type s, and to be equal to zero otherwise. $R_n(t; s) = (1/n)(Z_1 + \cdots + Z_n)$.

Let $V(v; y)$ denote the probability assigned by the density $f(y)$ to the sphere of volume v and center at y. In any closed region in which $f(y)$ is continuous, $V(v; y)$ can be written as $vf(y) + v\epsilon(y; v)$, where $\epsilon(y; v)$ approaches zero as v approaches zero, uniformly in y throughout the region. Clearly, $E\{Z_i\}$ is equal to

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{(n-1)!}{s!(n-1-s)!} \left[V\left(\frac{th(y)}{n}; y\right)\right]^s \cdot \left[1 - V\left(\frac{th(y)}{n}; y\right)\right]^{n-1-s} f(y)\, dy_1 \cdots dy_k. \tag{2.1}$$

The region of integration can be broken into a finite number of subregions, in the interior of each of which $f(y)$ and $h(y)$ are continuous. A closed subset of each such subregion can be found so that the measure of the set of points in k-dimensional space outside these closed subsets is arbitrarily small. Within each such closed subset, we may write

$$V\left(\frac{th(y)}{n}; y\right) = \frac{th(y)}{n} f(y) + \frac{th(y)}{n} \epsilon\left(y; \frac{th(y)}{n}\right),$$

where $\epsilon(y; th(y)/n)$ approaches zero as n increases, uniformly in y in the closed subset. Then it follows easily that the multiple integral (2.1) converges to $S(t; s)$ as n increases, so that $E\{R_n(t; s)\}$ approaches $S(t; s)$ as n increases.

To complete the proof that $R_n(t; s)$ converges stochastically to $S(t; s)$, we shall show that Var $\{R_n(t; s)\}$ approaches zero as n increases. Var $\{R_n(t; s)\}$ is equal to

$$\frac{1}{n^2} \sum_{i=1}^{n} \text{Var}\{Z_i\} + \frac{1}{n^2} \sum_{i \neq j} \sum \text{Cov}\{Z_i, Z_j\}.$$

Since $\{Z_i\}$ are uniformly bounded variables, $(1/n^2)\sum \text{Var}\{Z_i\}$ approaches zero as n increases. Therefore, to show that Var $\{R_n(t; s)\}$ approaches zero, it

will suffice to show that $E\{Z_i Z_j\}$ approaches $[S(t; s)]^2$ as n increases, since this implies that Cov $\{Z_i, Z_j\}$ approaches zero. $E\{Z_i Z_j\}$ is equal to

$$\int \underset{R_1}{\cdots} \int \frac{(n-2)!}{s!s!(n-2-2s)!}\left[V\left(\frac{th(y)}{n}; y\right)\right]^s \left[V\left(\frac{th(z)}{n}; z\right)\right]^s$$

$$(2.2) \qquad \cdot \left[1 - V\left(\frac{th(y)}{n}; y\right) - V\left(\frac{th(z)}{n}; z\right)\right]^{n-2-2s}$$

$$\cdot f(y)f(z)\, dy_1 \cdots dy_k\, dz_1 \cdots dz_k$$

$$+ \int \underset{R_2}{\cdots} \int Q(y, z) f(y) f(z)\, dy_1 \cdots dy_k\, dz_1 \cdots dz_k,$$

where R_1 is the region in (y, z) space such that the k-dimensional sphere of volume $th(y)/n$ centered at y does not intersect the k-dimensional sphere of volume $th(z)/n$ centered at z; R_2 is the remainder of (y, z) space; and $Q(y, z)$ is the conditional probability that the spheres around X_i and X_j are both of type s, given that $X_i = y$ and $X_j = z$. Clearly, the second integral in (2.2) approaches zero as n increases, and the first approaches $[S(t; s)]^2$. This completes the proof that $R_n(t; s)$ converges stochastically to $S(t; s)$ as n increases.

3. Application to multivariate tests of fit. Suppose the density of X, $f(y_1, \cdots, y_k)$ say, is unknown, and the hypothesis to be tested is that almost everywhere over a given region R, $f(y_1, \cdots, y_k) = g(y_1, \cdots, y_k)$, where $g(y_1, \cdots, y_k)$ is a given piecewise continuous function, $g(y_1, \cdots, y_k) \geqq B > 0$ at every point of R, and

$$1 \geqq \int \underset{R}{\cdots} \int g(y_1, \cdots, y_k)\, dy_1 \cdots dy_k > 0.$$

The hypothesis says nothing about $f(y_1, \cdots, y_k)$ outside the region R.

To construct a test of this hypothesis, we apply the result of Section 2 with $t = 1$, $s = 1$, and $h(y_1, \cdots, y_k) = 1/g(y_1, \cdots, y_k)$ for $(y_1, \cdots, y_k)$ in R, $h(y_1, \cdots, y_k) = 0$ elsewhere. Then

$$S(1; 1) = \int \underset{R}{\cdots} \int f(y) \frac{f(y)}{g(y)} \exp\left\{\frac{-f(y)}{g(y)}\right\} dy_1 \cdots dy_k$$

Using the fact that the function ue^{-u} takes on its absolute maximum at $u = 1$, we find that

$$S(1; 1) \leqq e^{-1} \int \underset{R}{\cdots} \int f(y)\, dy_1 \cdots dy_k,$$

with equality holding if and only if $g(y) = f(y)$ almost everywhere on R where $f(y) > 0$. Denote by $Q(n)$ the proportion of the observed points $X_1, X_2, \cdots, X_n$ that fall in the region R. $Q(n)$ converges stochastically to

$$\int \underset{R}{\cdots} \int f(y)\, dy_1 \cdots dy_k$$

as n increases. Thus, if the hypothesis is true, $R_n(1; 1)$ converges stochastically to

$$c^{-1} \int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k,$$

and $Q(n)$ converges stochastically to

$$\int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k.$$

Conversely, if $R_n(1; 1)$ converges stochastically to

$$c^{-1} \int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k$$

and $Q(n)$ converges stochastically to

$$\int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k,$$

then

$$S(1; 1) = c^{-1} \int \underset{R}{\cdots} \int f(y)\, dy_1 \cdots dy_k,$$

so the hypothesis is true.

For a given n, we define the following test T_n of the hypothesis. Accept the hypothesis if and only if

$$\left| R_n(1; 1) - c^{-1} \int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k \right| < A_n$$

and

$$\left| Q(n) - \int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k \right| < B_n,$$

where A_n, B_n are numbers chosen to give the desired level of significance, and it may (and will) be assumed that A_n and B_n both approach zero as n increases. From the discussion above, it is clear that the sequence of tests $\{T_n\}$ is consistent against any piecewise continuous alternative $f(y)$. To set the exact values of A_n, B_n the joint distribution of $Q(n)$ and $R_n(1; 1)$ would be required, but this distribution is unknown, although the author conjectures that it is asymptotically normal. However, given the function $g(y)$, the region R, and an alternative $f(y)$, the integrals (2.1) and (2.2) can be evaluated, at least approximately, and then Chebyshev inequalities can be used to give an upper bound to the level of significance and a lower bound to the power, for a given choice of A_n and B_n.

There are other consistent tests for the hypothesis under discussion: the chi-square test and obvious extensions of the univariate Kolmogorov-Smirnov

and von Mises tests. A comparison of the power functions of these tests would be of great interest. Almost nothing is known of the small-sample power of any of these tests. The large-sample power of the chi-square test is known. It is the author's conjecture that the limiting joint distribution of $Q(n)$ and $R_n(1; 1)$ is bivariate normal under the alternatives as well as under the hypothesis. If this conjecture could be proved, the asymptotic power of the proposed test would be known.

TABLES FOR OBTAINING NON-PARAMETRIC TOLERANCE LIMITS

By Paul N. Somerville[1]

General Analysis Corporation, Sierra Vista, Arizona

The general consideration of non-parametric tolerance limits had its origin with Wilks [10]. Wilks showed that for continuous populations, the distribution of P, the proportion of the population between two order statistics from a random sample, was independent of the population sampled, and was in fact a function only of the particular order statistics chosen. Wald [9] and Tukey [8] extended the method to multivariate populations, Tukey being responsible for the term "statistically equivalent block." Their work was extended further by Fraser [2], [3]. Murphy [4] presented graphs of minimum probable coverage by sample blocks determined by order statistics of a sample from a population with a continuous but unknown c.d.f. This note extends the results of Murphy, and tabularizes the results in a manner which eliminates or minimizes interpolation, particularly with respect to m, in a large number of cases. The form of Table I parallels the tables of Eisenhart, Hastay and Wallis [1] "Tolerance Factors for Normal Distributions."

Let P represent the proportion of the population between the r^{th} smallest and the s^{th} largest value in a random sample of n from a population having a continuous but unknown distribution function. Table I gives the largest value of $m = r + s$ such that we have confidence of at least that 100 P percent of the population lies between the r^{th} smallest and s^{th} largest in the sample. Note, that we may choose any $r, s \geqq 0$ such that $r + s = m$. We must, of course, decide upon the values of r and s independently of the observations in the sample. We obtain one-sided confidence intervals when we use $r = 0$ or $s = 0$ for a given m. The values of m are the largest such that

$$\gamma \leqq I_{1-P}(m, n - m + 1)$$

where I is the incomplete Beta function tabulated in [5] and [7].

Received March 21, 1957; revised January 15, 1958.

[1] Part of this work was done while the author was a guest worker at the National Bureau of Standards.

Table II gives the confidence γ that 100 P percent of the population lies between the largest and smallest of a random sample of n.

In the case where we are dealing with a multivariate population, we take m to be the number of blocks (See Tukey [8]) excluded from the tolerance region.

REFERENCES

[1] C E EISENHART, M. W. HASTAY, AND W. A. WALLIS, *Techniques of Statistical Analysis*, McGraw-Hill Book Company Inc., New York, 1947, Chapter 2
[2] D A. S FRASER, "Sequentially determined statistically equivalent blocks," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 372–381.
[3] D. A. S FRASER, "Nonparametric tolerance regions," *Ann. Math. Stat.*, Vol 24 (1953), pp. 44–55.
[4] R. B. MURPHY, "Nonparametric tolerance limits," *Ann Math. Stat.*, Vol 19 (1948), pp. 581–589.
[5] KARL PEARSON, *Tables of the Incomplete Beta-function*, Cambridge University Press (1934).
[6] HARRY G. ROMIG, "50–100 Binomial Tables," John Wiley and Sons, New York (1947).
[7] CATHERINE M. THOMPSON, "Tables of the Percentage Points of the Incomplete Beta-function," *Biometrika*, Vol. 32 (1941–42), pp. 151–181.
[8] J W. TUKEY, "Non-parametric estimation, II; Statistically equivalent blocks and tolerance regions, the continuous case," *Ann. Math. Stat*, Vol 18 (1947), pp. 529–539.
[9] A. WALD, "An extension of Wilks' method for setting tolerance limits," *Ann. Math. Stat*, Vol. 14 (1943), pp. 45–55.
[10] S. S WILKS, "Determination of sample sizes for setting tolerance limits," *Ann Math Stat.*, Vol. 12 (1941), pp. 91–96.
[11] Tables of the Cumulative Binomial Probability Distribution, *Annals of the Computation Laboratory of Harvard University*, Vol XXXV (1955)

NONPARAMETRIC ESTIMATION OF SAMPLE PERCENTAGE POINT STANDARD DEVIATION

BY JOHN E. WALSH

Military Operations Research Division, Lockheed Aircraft Corporation, Burbank, California

1. Summary. The available data consists of a random sample $x(1) < \cdots < x(n)$ from a reasonably well-behaved continuous statistical population. The problem is to estimate the standard deviation of a specified $x(r)$ that is not in the tails of the sample. The estimates examined are of the form

$$a[x(r + i) - x(r - i)]$$

and the explicit problem consists of determining suitable values for a and i. The solution

$$a = (\tfrac{1}{2})(n + 1)^{-3/10}\{[r/(n + 1)][1 - r/(n + 1)]\}^{1/2},\ i \doteq (n + 1)^{4/5}$$

Received November 15, 1957

The statistic $s[x(r)]$ has the smallest order error of all expected value estimates of $\sigma\{x(r)\}$ that are of the form $a[x(r+i) - x(r-i)]$, where i is $o(n)$. Here the order of the error of an estimate is considered to be the larger of the order of $E(\text{estimate}) - \sigma\{x(r)\}$ and the order of $\sigma\{\text{estimate}\}$.

The notation $f[x]$ is used to represent the probability density function of the population sampled. For the situation considered,

$$\sigma\{x(r)\} = \sqrt{p_r q_r}/\sqrt{n+1}f[\theta(p_r)] + O(n^{-3/2}).$$

This relation shows that a modification of $s[x(r)]$ can be used in estimating the value of the density function at the point $x = \theta(p_r)$. Explicitly,

$$\sqrt{n+1}s[x(r)]/\sqrt{p_r q_r}$$

furnishes an expected value estimate of $1/f[\theta(p_r)]$ that is accurate to terms of order $n^{-2/5}$.

3. Derivations. This section contains a verification of the properties stated for $s[x(r)]$ in the preceding sections.

Consider any integer t such that $1 \leqq t \leqq n$. From the results of [1],

$$(1) \qquad \begin{aligned} E[x(t)] &= \theta(p_t) - \frac{p_t q_t f'[\theta(p_t)]}{2(n+2)f[\theta(p_t)]^2} + O(n^{-2}), \\ \sigma\{x(t)\} &= \sqrt{p_t q_t}/\sqrt{n+1}f(\theta(p_t)] + O(n^{-3/2}). \end{aligned}$$

Here $p_t = t/(n+1)$ and $q_t = 1 - p_t$. These expansions, combined with appropriate use of Taylor series expansions, form the basis for the derivations.

Let $i = \epsilon(n+1)^\alpha = \text{integer}$, where $0 \leqq \alpha < 1$ and both ϵ and α are $0(1)$. Using (1) and expanding around r in Taylor series,

$$E[x(r+i)] = \theta(p_r) + \frac{\epsilon}{(n+1)^{1-\alpha}f[\theta(p_r)]} - \left[\frac{\epsilon^2}{(n+1)^{2-2\alpha}} + \frac{p_r q_r}{2(n+2)}\right]\frac{f'[\theta(p_r)]}{f[\theta(p_r)]^3} + O(n^{-3+3\alpha}) + O(n^{-2+\alpha})$$

$$E[x(r-i)] = \theta(p_r) - \frac{\epsilon}{(n+1)^{1-\alpha}f[\theta(p_r)]} - \left[\frac{\epsilon^2}{(n+1)^{2-2\alpha}} + \frac{p_r q_r}{2(n+2)}\right]\frac{f'[\theta(p_r)]}{f[\theta(p_r)]^3} + O(n^{-3+3\alpha}) + O(n^{-2+\alpha})$$

$$E\{a[x(r+i) - x(r-i)]\} = \frac{2a\epsilon}{(n+1)^{1-\alpha}f[\theta(p_r)]} + O(an^{-3+3\alpha}) + O(an^{-2+\alpha})$$

$$\sigma\{a[x(r+i) - x(r-i)]\} = a\sqrt{2\epsilon}/(n+1)^{1-\alpha/2}f[\theta(p_r)] + O(an^{-3/2+\alpha}).$$

The problem is to use these relations to determine suitable values for ϵ, α, and a.

Since $a[x(r+i) - x(r-i)]$ is an expected value estimate of $\sigma\{x(r)\}$,

$$2a\epsilon/(n+1)^{1-\alpha} = \sqrt{p_r q_r}/(n+1)^{1/2}, \quad \text{or} \quad a = (1/2\epsilon)\sqrt{p_r q_r}(n+1)^{1/2-\alpha}$$

Using this expression for a,

$$E\{a[x(r+i) - x(r-i)]\} = \sigma\{x(r)\} + O(n^{-5/2+2\alpha}) + O(n^{-3/2})$$
$$\sigma\{a[x(r+i) - x(r-i)]\} = O(n^{-1/2-\alpha/2}).$$

Thus increasing α decreases the order of magnitude of

$$\sigma\{a[x(r+i) - x(r-i)]\},$$

but increases the order of

$$E\{a[x(r+i) - x(r-i)]\} - \sigma\{x(r)\}.$$

Hence the order of the error is minimized when

$$-1/2 - \alpha/2 = -5/2 + 2\alpha.$$

Thus $\alpha = 4/5$ appears to be the most desirable choice for α.

In $\sigma\{a[x(r+i) - x(r-i)]\}$, the parameter ϵ appears predominantly as the factor $1/\sqrt{\epsilon}$. In $E\{a[x(r+i) - x(r-i)]\} - \sigma\{x(r)\}$ the predominant factor is ϵ^2. Solution of the equation

$$\epsilon^2 = 1/\sqrt{\epsilon}$$

suggests that $\epsilon = 1$ is an appropriate compromise choice for ϵ.

Use of $\alpha = 4/5$, $\epsilon = 1$, and the expression for a yields the results

$$i = (n+1)^{4/5}, \qquad a = \tfrac{1}{2}(n+1)^{-3/10}\sqrt{p_r q_r},$$

and verifies the properties stated for $s[x(r)]$.

REFERENCE

[1] F. N. David and N. L. Johnson, "Statistical treatment of censored data. Part I. Fundamental formulae," *Biometrika*, Vol. 41 (1954), p. 228–240.

A UNIQUENESS PROPERTY NOT ENJOYED BY THE NORMAL DISTRIBUTION

By George P. Steck

Sandia Corporation

1. Summary. It is well known that if X and Y (or $1/X$ and $1/Y$) are independently normally distributed with mean zero and variance σ^2, then X/Y has a Cauchy distribution. It is the purpose of this note to show that the converse statement is not true. That is, the fact that the ratio of two independent, identically distributed, random variables X and Y follows a Cauchy distribution is not sufficient to imply that X and Y (or $1/X$ and $1/Y$) are normally distributed. This will be shown by exhibiting several counterexamples.

Received October 21, 1957; revised December 26, 1957.

2. Construction of counterexamples. Let X and Y be independent, identically distributed, random variables with common symmetric density function p. Let φ denote the characteristic function of $(4/\pi)\log|X|$, let $Z = X/Y$, and let ω denote the characteristic function of $(4/\pi)\log|Z|$. The fact that Z has a Cauchy distribution implies that

$$p_{4/\pi \log|Z|}(u) = \frac{1}{4\cosh \pi u/4}, \qquad -\infty < u < +\infty,$$

and, hence,

$$\omega(t) = \frac{1}{4}\int_{-\infty}^{\infty} \frac{e^{itu}\,du}{\cosh \pi u/4} = \frac{1}{\cosh 2t}.$$

Since

$$\frac{4}{\pi}\log|Z| = \frac{4}{\pi}\log|X| - \frac{4}{\pi}\log|Y|,$$

it follows that

$$\varphi(t)\cdot\varphi(-t) = \frac{1}{\cosh 2t},$$

and, therefore,

$$\text{(1)} \qquad \varphi(t) = \frac{e^{i\theta(t)}}{(\cosh 2t)^{\frac{1}{2}}} = \frac{\cos\theta(t)}{(\cosh 2t)^{\frac{1}{2}}} + i\frac{\sin\theta(t)}{(\cosh 2t)^{\frac{1}{2}}},$$

where θ is continuous, real, odd, and of such a form that φ is a characteristic function.

Since $\varphi(t)$ must be inverted by contour integration to find corresponding density functions, equation (1) suggests that θ be chosen so as to eliminate the square root. The relations

$$\cosh 2t = \cosh^2 t + \sinh^2 t = 1 + 2\sinh^2 t$$

provide two functions θ which accomplish this, namely,

$$\theta_1(t) = \arctan \tanh t$$

and

$$\theta_2(t) = \arctan \sqrt{2}\sinh t.$$

Other functions θ which immediately suggest themselves, even though they do not eliminate the square root, are

$$\theta_3(t) \equiv 0, \quad \text{and} \quad \theta_4(t) = t.$$

The corresponding functions φ are

$$\varphi_1(t) = \frac{\cosh t}{\cosh 2t} + i\frac{\sinh t}{\cosh 2t},$$

$$\varphi_2(t) = \frac{1}{\cosh 2t} + i\sqrt{2}\,\frac{\sinh t}{\cosh 2t},$$

$$\varphi_3(t) = \frac{1}{(\cosh 2t)^{\frac{1}{2}}} \quad \text{and} \quad \varphi_4(t) = \frac{e^{it}}{(\cosh 2t)^{\frac{1}{2}}}.$$

If these functions φ are inverted (see [1], pp. 388–389, and [2], p. 30) and a change of variable made from $(4/\pi)\log|X|$ to X, assuming X symmetric, then the corresponding density functions are

$$p_1(x) = \frac{\sqrt{2}}{\pi}\,\frac{x^2}{1+x^4}, \qquad -\infty < x < +\infty,$$

$$p_2(x) = \frac{2}{\pi}\,\frac{x^4}{(1+x^2)(1+x^4)}, \qquad -\infty < x < +\infty,$$

$$p_3(x) = \frac{1}{2\pi^2\sqrt{2\pi}\,|x|}\left|\Gamma\left(\frac{1}{4} + \frac{i\log|x|}{\pi}\right)\right|^2, \qquad -\infty < x < +\infty,$$

and

$$p_4(x) = \frac{1}{2\pi^2\sqrt{2\pi}\,|x|}\left|\Gamma\left(\frac{1-i}{4} + \frac{i\log|x|}{\pi}\right)\right|^2, \qquad -\infty < x < +\infty.$$

Using $\theta(-t)$ instead of $\theta(t)$ provides additional densities $p^*(x) = p(1/x)/x^2$ (if p is the density function of X then p^* is the density function of $1/X$). For example,

$$p_1^*(x) = \frac{\sqrt{2}}{\pi}\,\frac{1}{1+x^4}, \qquad -\infty < x < +\infty,$$

and

$$p_2^*(x) = \frac{2}{\pi}\,\frac{1}{(1+x^2)(1+x^4)}, \qquad -\infty < x < +\infty.$$

REFERENCES

[1] J. D. Bierens de Haan, *Nouvelles tables d'Intégrals Définies*, G. E. Stechert, 1939.
[2] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of integral transforms*, Vol. 1, McGraw-Hill, 1954.

ESTIMATION OF A REGRESSION LINE WITH BOTH VARIABLES SUBJECT TO ERROR UNDER AN UNUSUAL IDENTIFICATION CONDITION

By Herman Rubin

University of Oregon

Suppose the random variables $w_j = (\xi_j, u_j, v_j)$ are independently and identically distributed with joint distribution F. Then if $\iiint e^{\alpha u+\beta v}\,dF(\xi, u, v)$ exists

Received December 13, 1956; revised February 10, 1958.

for all α, β in a neighborhood of 0 and $\iiint e^{t\xi}\, dF(\xi, u, v)$ does not exist for all t in any neighborhood of 0, Jeeves [1] has shown that the parameter θ in

$$x_j = \xi_j \cos\theta + u_j, \qquad y_j = \xi_j \sin\theta + v_j, \tag{1}$$

is identified. We shall construct a consistent estimate of θ (mod π) if these conditions are satisfied

First, let us consider a univariate distribution G with moment generating function g. Then $g(t) = \sum (\mu_n / n!)\, t^n$, μ_n the nth moment, and if the radius of convergence is r, it is well known that

$$\rho = \frac{1}{r} = \overline{\lim}\left(\frac{|\mu_n|}{n!}\right)^{1/n} \tag{2}$$

As easy application of Liapounoff's inequality and Stirling's formula shows that

$$\rho = \overline{\lim}\,\frac{(\mu_{2n})^{1/2ne}}{2n} \tag{3}$$

Therefore a natural procedure would seem to be to consider the sample moments $m_{2n}(\phi)$ of $x_j \sin\phi - y_j \cos\phi$ and to define $\hat{\theta}$ as that value of ϕ minimizing $\max_n\, (m_{2n}(\phi))^{1/2n}/n$. For fixed sample size, this maximum exists. We shall show that this estimate is indeed consistent, and even converges with probability one to θ.

First let us show that $\max_n m_{2n}(\theta)^{1/2n}/n$ is bounded as a function of the sample size N with probability one. Let

$$\psi(t) = E(\cosh[t(u_j \sin\theta - v_j \cos\theta)]) = \sum \frac{\mu_{2n}\, t^{2n}}{(2n)!},$$

where μ_{2n} is the $2n$th moment of $u_j \sin\theta - v_j \cos\theta$. Then [2], for $|t| \leqq s < r$, $\psi_N(t) = \sum m_{2n} t^{2n}/(2n)!$ converges to $\psi(t)$ uniformly with probability one. Thus $\psi_N(t)$ is bounded with probability one, and since $m_{2n}{}^{1/2n}/n \leqq (K/t)\,[\psi_N(t)]^{1/2n}$, $\max_n m_{2n}(\theta)^{1/2n}/n$ is bounded with probability one. Hence with probability greater than $1 - \epsilon/3$, H_ϵ can be used for the bound. Similarly,

$$\max_{n,\phi} \frac{\left(\frac{1}{N}\sum(u_j \sin\phi - v_j \cos\phi)^{2n}\right)^{1/2n}}{n} < K_\epsilon$$

for all N with probability greater than $1 - \epsilon/3$. Let δ be given, $0 < \delta < \pi$, and let γ_n be the nth moment of ξ_j. Since $(\gamma_{2n})^{1/2n}/n$ can be made arbitrarily large by selecting n large enough, select n so that

$$\frac{(\gamma_{2n})^{1/2n}}{n} > \frac{H_\epsilon + K_\epsilon}{\sin\delta}.$$

Then with probability greater than $1 - \epsilon/3$,

$$\frac{(1/N \sum \xi_j^{2n})^{1/2n}}{n} > \frac{H_\epsilon + K_\epsilon}{\sin\delta}$$

for all N sufficiently large. By Minkowski's inequality

$$m_{2n}(\phi)^{1/2n} \geqq |\sin(\phi - \theta)| \left(\frac{1}{N}\sum \xi_j^{2n}\right)^{1/2n} - \left(\frac{1}{N}\sum (u_j \sin\phi - v_j \cos\phi)^{2n}\right)^{1/2n}$$

Therefore with probability greater than $1 - \epsilon$,

$$\frac{\max_n (m_{2n}(\phi))^{1/2n}}{n} > \frac{\max_n (m_{2n}(\theta))^{1/2n}}{n}$$

for all N sufficiently large for all ϕ not in the interval $(\theta - \delta, \theta + \delta)$ (mod

REFERENCES

[1] T. A. Jeeves, "Identifiability and almost sure estimability of linear structure in dimensions," University of California, 1952, unpublished paper.
[2] Herman Rubin, "Uniform convergence of random functions with applications to sta tics," *Ann. Math. Stat.*, vol. 27 (1956), pp.200–203.

ON THE DECOMPOSITION OF CERTAIN χ^2 VARIABLES

By Robert V. Hogg and Allen T. Craig

University of Iowa

It is well known that if the sum, say $Q = Q_1 + Q_2$, of two stochastica independent variables is χ^2 with r d.f., and if Q_1 is also χ^2 with r_1 d.f., then is likewise χ^2 with $r_2 = r - r_1$ d.f. If the hypothesis of stochastic independe is removed, little can be said about Q_2. It seems to us quite interesting that the variables under consideration are real symmetric quadratic forms in eit central or non-central, stochastically independent or dependent normal variab and if the hypothesis of stochastic independence of Q_1 and Q_2 is replaced by weaker hypothesis $Q_2 \geqq 0$, then Q_1 and Q_2 are stochastically independent so t Q_2 is itself a χ^2 variable with $r_2 = r - r_1$ d.f.

Before we state our theorem, we recall [1] that the real symmetric quadr form $Y'BY$ in n mutually stochastically independent normal variables Y' $(y_1, y_2, \cdots, y_n)$ with unit variances and means $U' = (u_1, u_2, \cdots, u_n)$ ha non-central χ^2 distribution whose characteristic function is

$$\varphi(t) = \exp\left[\frac{it\theta}{1 - 2it}\right] \Big/ (1 - 2it)^{r/2}$$

if and only if $B^2 = B$. Here, $\theta = U'BU$ and r is the rank of B.

Theorem. *Let $Q = Q_1 + \cdots + Q_{k-1} + Q_k$, where $Q = X'AX$ and Q_j $X'A_jX$, $j = 1, 2, \cdots, k$, are real symmetric quadratic forms in n normally tributed variables $X' = (x_1, x_2, \cdots, x_n)$ with means $M' = (m_1, m_2, \cdots,$ and real symmetric definite positive variance-covariance matrix V. Let $Q, Q_1, \cdots$*

Received February 25, 1957; revised November 27, 1957.

Q_{k-1} have non-central χ^2 distributions with parameters r, θ and r_j, θ_j, ($j = 1, \cdots, k-1$), respectively and let Q_k be non-negative. Then $Q_1, Q_2, \cdots, Q_k$ are mutually stochastically independent and Q_k has a non-central χ^2 distribution with parameters $r_k = r - \sum_1^{k-1} r_j$, $\theta_k = \theta - \sum_1^{k-1} \theta_j$.

Proof. We first prove the theorem for $k = 2$. There exists a real symmetric positive definite matrix C such that $C'C = V$. If we let $X = CY$, $Y' = (y_1, y_2, \cdots, y_n)$, and at the same time let $M = CU$, $U' = (u_1, u_2, \cdots, u_n)$, then $y_1, y_2, \cdots, y_n$ are mutually stochastically independent normal variables with unit variances and means $U' = (u_1, u_2, \cdots, u_n)$. Also

$$X'AX = X'A_1X + X'A_2X$$

becomes $Y'BY = Y'B_1Y + Y'B_2Y$, where $B = C'AC$, $B_1 = C'A_1C$, $B_2 = C'A_2C$, and $B = B_1 + B_2$. By hypothesis, $Y'BY$ and $Y'B_1Y$ have *non-central* χ^2 distributions and $Y'B_2Y \geqq 0$. Thus $B^2 = B$ and $B_1^2 = B_1$. With a suitably chosen orthogonal matrix L, $L'BL$ is a diagonal matrix having r ones and $n - r$ zeros on the principal diagonal. Since B_1 and B_2 are semi-definite positive, each element on the principal diagonal of $L'B_1L$ and $L'B_2L$ is non-negative and hence each of these matrices has a zero on the principal diagonal corresponding to each zero on that of $L'BL$. Moreover all elements in the rows and columns of $L'B_1L$ and $L'B_2L$ in which these zeros appear are likewise zero. If we properly choose our notation we may the write $L'BL = L'B_1L + L'B_2L$, using submatrices, as

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} G_r & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} H_r & 0 \\ 0 & 0 \end{pmatrix}.$$

If we multiply on the left by $L'B_1L$ and make use of $B_1^2 = B_1$, we have

$$L'B_1B_2L = 0.$$

That is, $B_1B_2 = 0$, so, by a result of Carpenter [1]. $Y'B_1Y$ and $Y'B_2Y$ (that is, Q_1 and Q_2) are stochastically independent. Since Q and Q_1 have non-central χ^2 distributions it follows that Q_2 has a non-central χ^2 distribution with parameters $r_2 = r - r_1$, $\theta_2 = \theta - \theta_1$. For $k > 2$, the proof of the theorem is easily completed by induction.

As an example, let $(x_1, y_1), \cdots, (x_n, y_n)$ denote a random sample from a bivariate normal distribution having unit variances, means m_x and m_y, and correlation coefficient ρ. It is fairly obvious that the left member and the first term of the right member of

$$\sum_1^n (x_i^2 - 2\rho x_i y_i + y_i^2)/(1 - \rho^2) = (n\bar{x}^2 - 2\rho n\bar{x}\bar{y} + n\bar{y}^2)/(1 - \rho^2)$$

$$+ \sum_1^n [(x_i - \bar{x})^2 - 2\rho(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2]/(1 - \rho^2)$$

have non-central χ^2 distributions with parameters $r = 2n$,

$$\theta = n(m_x^2 - 2\rho m_x m_y + m_y^2)/(1 - \rho^2)$$

and $r_1 = 2$, $\theta_1 = n(m_x^2 - 2\rho m_x m_y + m_y^2)/(1 - \rho^2)$ respectively. Accordingly, the non-negative form

$$\sum_1^n [(x_i - \bar{x})^2 - 2\rho(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2]/(1 - \rho^2)$$

has a central χ^2 distribution with $2n - 2$ degrees of freedom.

REFERENCE

[1] OSMER CARPENTER, "Note on the extension of Craig's theorem to noncentral variates," *Ann. Math. Stat.*, Vol. 21 (1950), p. 455.

A NOTE ON THE GENERATION OF RANDOM NORMAL DEVIATES[1]

BY G. E. P. BOX AND MERVIN E. MULLER

Princeton University

1. Introduction. Sampling experiments often require the generation of large numbers of random normal deviates. When an electronic computer is used it is desirable to arrange for the generation of such normal deviates within the machine itself rather than to rely on tables. Pseudo random numbers can be generated by a variety of methods within the machine and the purpose of this note is to give what is believed to be a new method for generating normal deviates from independent random numbers. This approach can be used on small as well as large scale computers. A detailed comparison of the utility of this approach with other known methods (such as: (1) the inverse Gaussian function of the uniform deviates, (2) Teichroew's approach, (3) a rational approximation such as that developed by Hastings, (4) the sum of a fixed number of uniform deviates and (5) rejection-type approach), has been made elsewhere [1] by one of the authors (M.M.). It is shown that the present approach not only gives higher accuracy than previous methods but also compares in speed very favourably with other methods.

2. Method. The following approach may be used to generate a pair of random deviates from the same normal distribution starting from a pair of random numbers.

Method: Let U_1, U_2 be independent random variables from the same rectangular density function on the interval (0, 1). Consider the random variables:

$$X_1 = (-2 \log_e U_1)^{1/2} \cos 2\pi U_2 \tag{1}$$
$$X_2 = (-2 \log_e U_1)^{1/2} \sin 2\pi U_2$$

Received October 30, 1957; revised January 31, 1958.

[1] Prepared in connection with research sponsored by the Office of Ordnance Research U. S. Army; Statistical Techniques Research Group, Princeton University, Contract No DA 36-034-ORD 2297.

Then (X_1, X_2) will be a pair of independent random variables from the same normal distribution with mean zero, and unit variance.

Justification: From (1) (giving attention to principal values), one obtains at once the inverse relationships:

$$U_1 = e^{\frac{-(X_1^2 + X_2^2)}{2}}$$

$$U_2 = -\frac{1}{2\pi} \arctan \frac{X_2}{X_1}.$$

It follows that the joint density of X_1, X_2 is

$$f(X_1, X_2) = \frac{1}{2\pi} e^{\frac{-(X_1 + X_2)}{2}} = \frac{1}{\sqrt{2\pi}} e^{\frac{-X_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-X_2^2}{2}} = f(X_1)f(X_2);$$

thus the desired conclusions, including the independence of X_1 and X_2 is obtained.

The above approach is motivated by the following considerations: the probability density of $f(X_1, X_2)$ is constant on circles, so $\Theta = \arctan X_2/X_1$ is uniformly distributed $(0, 2\pi)$. Further, the square of the length of the radius vector $r^2 = X_1^2 + X_2^2$ has a Chi-squared distribution with two degrees of freedom. If U has a rectangular density on (0, 1) then $-2 \log_e U$ has a Chi-squared distribution with two degrees of freedom. Proceeding in the reverse order we arrive at (1).

3. Generalizations and other random variables. Observations from the Chi-squared distribution with $2k$ degrees of freedom can of course be generated by adding together the k terms, $\sum_{i=1}^{k} (-2 \log_e U_i)$ and for Chi-squared with $2k + 1$ degrees of freedom one may add the square of a normal deviate generated by the above method. Deviates from the F-distribution and for the t-distribution are obtained by calculating the appropriate ratio of deviates generated as above. From independent random normal deviates well known methods can of course be used to generate n-dimensional normal deviates with arbitrary means and variance-covariance matrix.

4. Convenience and accuracy. The method suggested here grew out of the desire to have a way of generating normal deviates which would be reliable in the tails of the distribution. Since most computing centers have library programs to compute values of trigonometric functions, logarithms, and square roots this approach requires little additional machine program writing. The accuracy obtained depends essentially on the precision of the available library programs, whereas that of other methods cannot readily be increased.

REFERENCE

[1] M. E. Muller, "Generation of Normal Deviates," Technical Report No. 13, Statistical Techniques Research Group, Department of Mathematics, Princeton University.

ON A THEOREM IN METRIC SPACES

By V. S. Varadarajan and R. Ranga Rao

Indian Statistical Institute

0. Introduction. In his paper "On a class of probability spaces" ([1]), D. Blackwell observed that the class of Borel sets of a metric space may be a separable σ-field without the metric space being separable. However, in a subsequent letter to one of the authors, he stated that the question remained open. The object of the present note is to prove that the separability of the σ-field of Borel sets implies separability of the metric space, assuming the continuum hypothesis. What is actually used, is not the continuum hypothesis but the following proposition, which we will abbreviate as $\mathcal{P}$: If u is an uncountable cardinal, $2^u > c$ (the cardinal of the continuum). This is easily deduced from the continuum hypothesis and it seems to us that it has not so far been proved without the continuum hypothesis (cf. [4]). The main conclusion is as follows: A metric space is separable if and only if the cardinality of the Borel sets is $\leqq c$, provided we assume $\mathcal{P}$. It is also shown that the above theorem implies $\mathcal{P}$.

1. The main result. We introduce certain notations. X is a metric space and $\mathcal{B}$ is the σ-field generated by open subsets of X. Sets of $\mathcal{B}$ are called Borel sets of X. $\mathcal{B}$ is called separable if there is a sequence $\{A_n\}$ of sets of $\mathcal{B}$ generating it. In that case, cardinality of $\mathcal{B}$ is $\leqq c$ ([2]). Before proving the main result we prove an auxiliary result, interesting in itself.

Theorem 1. *X is separable if and only if every disjoint family of nonempty open subsets of X is countable.*

Proof. If X is separable, its topology has a countable basis $G_1, G_2, \cdots$. Since any nonempty open set of X contains a nonempty G_i, the existence of an uncountable disjoint family of nonempty open subsets of X implies the existence of an uncountable disjoint family of nonempty G_i's, which is impossible. To prove the converse, let us suppose that every disjoint family of nonempty open subsets of X is countable. Let n be an integer $\geqq 1$ and let $\mathcal{K}_n$ be defined as fol lows: $\mathcal{K}_n = \{A: A \subset X; x, y \in A \Rightarrow d(x, y) > 1/n\}$. Elements of $\mathcal{K}_n$ are sub sets of X and are partially ordered by the relation of set inclusion. Furthe every linearly ordered sub-family of $\mathcal{K}_n$ has a supremum in the family (namel the set-union) and hence, by Zorn's lemma, there are maximal elements contai ing any element of $\mathcal{K}_n$, in particular any point of X. Let A_n be one such no empty maximal element. Maximality of A_n implies that if $y \in X - A_n$, $d(y, x)$ $1/n$ for some $x \in A_n$. Further, each A_n must be countable, as otherwise, tl spheres with centres at the points of A_n and radii $(1/2n)$ will be an uncounta disjoint family of nonempty open subsets of X.

Let now n run over $1, 2, \cdots$ and set $A = \bigcup_n A_n$. A is countable and for a $y \in X - A$ and any positive integer n, there is an $x_n \in A_n$ such that $d(y, x_n)$

Received November 8, 1957.

$1/n$. This shows that A is dense in X. Since A is countable, this completes the proof that X is separable.

Remark. This result need not be true if X is not metric. See, for example, [3].

We now prove our main result.

THEOREM 2. *(Under assumption $\mathcal{P}$) X is separable if and only if cardinality of $\mathcal{B} \leqq c$. In particular, X is separable if and only if $\mathcal{B}$ is separable.*

Proof. If X is separable, its topology has a countable base $G_1, G_2, \cdots$ which generates $\mathcal{B}$ and hence cardinality of $\mathcal{B} \leqq c$. Conversely let cardinality of $\mathcal{B}$ be $\leqq c$. If $\{A_\alpha\}_{\alpha \in I}$ is any disjoint family of nonempty open subsets of X, then, every subunion of the A_α's is open and hence $\varepsilon\ \mathcal{B}$. There are 2^u such subunions where u is the cardinal of I and since cardinality of $\mathcal{B}$ is $\leqq c$, we have $2^u \leqq c$. This however implies (in virtue of assumption $\mathcal{P}$) that $u \leqq \aleph_0$. Theorem 1 now applies and proves that X is separable. This completes the proof.

Remark. We can show that Theorem 2 implies $\mathcal{P}$. To see this, let X be an uncountable set with cardinal u. Give X the discrete topology so that $\mathcal{B}$ is the class of all subsets of X. Cardinality of $\mathcal{B}$ is thus 2^u and since X is not separable, Theorem 2 implies that $2^u > c$. This is precisely assumption $\mathcal{P}$

REFERENCES

[1] D. BLACKWELL, "On a class of probability spaces, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol I, University of California Press, 1956, p 2

[2] P. R. HALMOS, *Measure Theory*, D Van Nostrand, 1950, p 26, 9(c)

[3] J. L KELLEY, *General Topology*, D Van Nostrand, 1955, p. 166, ex 0, (f)

[4] W SIERPINSKI, "Sur l'hypothèse du continuum " *Fundamenta Mathematica*, Vol 5, p. 182.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Ames, Iowa Meeting of the Institute, April 3-5, 1958.)

11. Bias and Confidence in Not-quite Large Samples. (Preliminary Report) JOHN W. TUKEY, Princeton University, (By Title).

The linear combination of estimates based on all the data with estimates based on parts thereof seems to have been first treated in print as a means of reducing bias by Jones (*J. Amer. Stat. Assn.*, Vol. 51 (1956), pp. 54-83). Let $y_{(.)}$ be the estimate based on all the data, $y_{(i)}$ that based on all but the ith piece, $\bar{y}_{(i)}$ the average of the $y_{(i)}$. Quenouille (*Biometrika*, Vol. 43 (1956), pp. 353-360) has pointed out some of the advantages of $ny_{(.)} - (n-1)\bar{y}_{(i)}$ as such an estimate of much reduced bias. Actually, the individual expressions $ny_{(.)} - (n-1)y_{(i)}$ may, to a good approximation, be treated as though they were n independent estimates. Not only is each nearly unbiased, but their average sum of squares of deviations is nearly $n(n-1)$ times the variance of their mean, etc. In a wide class of situations they behave rather like projections from a non-linear situation on to a tangent linear situation. They may thus be used in connection with standard confidence procedures to set closely approximate confidence limits on the estimand. (Received December 26, 1957.)

12. Limiting Distributions of k-sample Test Criteria of Kolmogorov-Smirnov-v. Mises Type. J. KIEFER, Cornell University, (By Title).

Let S_j be the sample d.f. of n_j independent, indentically distributed random variable: with common unknown continuous d.f. F_j $(1 \leqq j \leqq k)$, the S_j being independent. For testing the hypothesis $H:F_1 = \cdots = F_k$, several criteria were suggested by the author in *Ann Math. Stat.*, Vol. 26 (1955), p. 775. Among these are $T = \sup_x \sum_j n_j[S_j(x) - \bar{S}(x)]^2$ and

$$W = \int \sum_j n_j[S_j(x) - \bar{S}(x)]^2 dS^*(x),$$

where $\bar{S} = \sum_j n_j S_j / \sum n_j$ and $S^* = \sum_j a_j S_j$ with $\sum_j a_j = 1$. It is proved by the metho indicated in the above reference that, under H, the limit of $P\{T < a^2\}$ as all $n_j \to \infty$, is

$$\frac{2^{(5-2k)/2}a^{1-k}}{\Gamma((k-1)/2)} \sum_{n=1}^{\infty} \frac{\alpha_n^{k-3} \exp[-\alpha_n^2/2a^2]}{[J_{(k-1)/2}(\alpha_n)]^2},$$

where $a > 0$ and α_n is the nth positive zero of the Bessel function $J_{(k-3)/2}$; alternative e pressions are also given. When $k = 4$ or $k = 2$ the summand above reduces to an elementa function; the latter case gives the Kolmogorov-Smirnov distribution, since $T^{1/2}$ is t Smirnov statistic when $k = 2$. The limiting d.f. of W is expressible in a series involvi Hermite polynomials when k is odd and Bessel functions when k is even. For $k = 2$, W the test suggested by Lehmann and Rosenblatt, and the above d.f. is the limiting ω^2 d in the form given by Anderson and Darling. (Received January 6, 1958.)

13. A Rule for Action Based on Percentage Changes in the Sample Mean. B. OWEN, Sandia Corporation, (By Title).

A random selection is made of n items from a normal population X, each item is measur once, and the sample mean $\bar{x}$ is computed. The sample items are identified by some me and the sample and the remaining population are mixed at random. They are then su jected to some condition, such as storage, after which the same items that were first sa

pled are measured again, giving a new mean $\bar{y}$ Some action is taken only if the new mean $\bar{y}$ differs from the old mean $\bar{x}$ by more than $p\bar{x}$, where $p > 0$. The probability of taking action using the above rule is shown to be expressible in terms of the bivariate normal cumulative probability function. If there is no change in the mean (from X to Y), the probability for action increases monotonically with an increase in the standard deviation If there is no increase in the standard deviation (from X to Y), the probability for action increases monotonically with any increase in the mean. However, for a fixed (relatively large) increase in the mean, the probability for action drops with increasing standard deviation and then increases (Received February 5, 1958)

14. On a Multivariate Gamma Distribution. P. R. KRISHNAIAH and M. M. RAO, University of Minnesota.

From the relation between the univariate Gamma and Gaussian distributions one naturally considers the corresponding p-variate cases Some writers in the past implied this in their approach to this problem. But the properties of the latter distribution are not well utilized. The derivation of a p-variate Gamma distribution by Krishnamoorthy and Parthasarathy (these *Annals*, 1951) and later used by Gurland (these *Annals*, 1955) was not too direct. In this paper a simple derivation using the too familiar properties of the Normal and Wishart distributions is given. Also some interesting connections between the Gamma and Gaussian distributions are discussed A special case for $p = 2$ (for correlations) has been given in Cramér's book (p. 317). But that property is shown to be true for $p > 2$ Next the "Arithmetical Character" of this distribution in the sense of P Lévy (*Proc Cambridge Philos. Soc.* (1948), p. 295) is considered Some small and large sample properties are also discussed. (Received February 6, 1958)

15. An Expression for the Cumulative Distribution Function of the Noncentral t-Distribution. D. B. OWEN, Sandia Corporation, (By Title)

The cumulative distribution function of the noncentral t-distribution may be expressed in terms of the univariate normal integral and elementary functions for an even number of degrees of freedom and for an odd number of degrees of freedom in terms of the univariate normal integral, elementary functions, and the $T(h, a)$ function. The $T(h, a)$ function was tabulated by the present author in *Ann. Math. Stat* , Vol 27 (1956), pp 1075–1090 The above results were obtained by repeated integration by parts For example, with one degree of freedom $\Pr(T \leqq t) = G(-\delta/\sqrt{1+t^2}) + 2T(\delta/\sqrt{1+t^2}, t)$, where $G(x)$ is the univariate normal integral from minus infinity to x, and δ is the noncentrality parameter. This expression is especially useful since the noncentral t-distribution has not been tabulated for one degree of freedom. (Received February 6, 1958)

16. The Fourth Product Moment of a Binary Random Process. J. A. McFADDEN, Purdue University, (introduced by Judah Rosenblatt) (By Title).

Let $x(t)$ describe a stationary random process, and let $y(t) = 1$ when $x(t) \geqq 0$ and $y(t) = -1$ when $x(t) < 0$. Let $s(\tau_1, \tau_2, \tau_3)$ denote the fourth product moment,

$$E[y(t)y(t+\tau_1)y(t+\tau_2)y(t+\tau_3)],$$

where $0 \leqq \tau_1 \leqq \tau_2 \leqq \tau_3$. If $x(t)$ is a Gaussian process, then s is related to the quadrivariate normal integral, which apparently cannot be expressed in closed form For practical applications it seems advisable to make different assumptions about $x(t)$ (or about $y(t)$. Let $E[y(t)] = 0$ and let all product moments of odd degree in $y(t)$ be zero Consider furthermore the zeros of the function $x(t)$. If the zeros obey the Poisson distribution, then a par-

y simple result follows for s and for all higher moments. Another assumption is the ng: Let unspecified events occur at times $t_1, t_2, \cdots$, according to the Poisson dis- on, the average number of events per unit time being denoted by α. If alternate (those at $t_1, t_3, \cdots$) are designated as zeros of $x(t)$, then the autocorrelation func- $y(t)$ is $E[y(t)y(t+\tau)] = e^{-\alpha\tau}\cos\alpha\tau$, and the desired fourth moment is

$$s = e^{-(u+w)}\cos u \cos w - e^{-(u+2v+w)}\sin u \sin w,$$

$u = \alpha\tau_1$, $v = \alpha(\tau_2 - \tau_1)$, and $w = \alpha(\tau_3 - \tau_2)$. (Received February 10, 1958.)

pproximate Solutions for the Probability Density of Zero-Crossing Intervals in a Gaussian Process. J. A. McFadden, Purdue University, (introduced by Judah Rosenblatt).

$x(t)$ be a stationary Gaussian process, and let $P_0(\tau)$ be the probability density of gths of intervals between successive zeros in this process. Under the assumption that gth of a given zero-crossing interval is independent of the sum of the previous 2) interval lengths, where m takes on *all* values, $m = 0, 1, 2, \cdots$, then the following l equation can be derived for $P_0(\tau)$: $P_0(\tau) = Q_1(\tau) - \int_0^\tau Q_2(l)P_0(\tau - l)\,dl$, where r is the conditional probability of a zero with *negative* slope in the interval $(t+\tau,$ $d\tau)$, given a zero with positive slope at time t, and $Q_2(\tau)\,d\tau$ is the conditional prob- of a zero with *positive* slope in $(t+\tau, t+\tau+d\tau)$, given a zero with positive slope ing expressions for $Q_1(\tau)$ and $Q_2(\tau)$ given by S. O. Rice, the integral equation has lved numerically for several choices of spectral density. The results compare favor- th experiment, and the agreement is much better than can be obtained by the usual l methods, i.e., assuming that all interval lengths are independent. (Received Febru- 1958.)

inimal Complete Classes of Tests. D. L. Burkholder, University of Illinois.

mal complete classes of tests are found for a number of common testing problems ng, for example, those listed by Lehmann and Scheffé in *Sankhyā*, Vol. 15 (1955), with respect to the exponential family of distributions. The proofs are based partly heory of complete and sufficient statistics and partly on other ideas needed and de- for those cases in which the hypothesis set ω and the alternative set $\Omega - \omega$ are ed by an indifference zone. The kinds of results obtained are illustrated in the fol- special case: Let X_1 and X_2 be independent random variables where X_i is binomial $0 < p_i < 1$, $i = 1, 2$. Let ω be a subset of $\{(p_1, p_2) \mid p_1 \geqq p_2\}$, $\Omega - \omega$ be a subset $p_2) \mid p_1 < p_2\}$, and suppose there are positive numbers m and M such that if

$$S = \{(p_1, p_2) \mid m < p_1/p_2 < M\}$$

ch of $S \cap \omega$ and $S \cap (\Omega - \omega)$ has the origin as a limit point. Let C be the class of φ of the form: $\varphi(x_1, x_2) = 1$ if $x_1 < c(x_1 + x_2)$, $= a(x_1 + x_2)$ if equality holds, $= 0$ se. Then C is the minimal complete class of tests for the problem of testing $(p_1, p_2)\ \varepsilon$ st $(p_1, p_2)\ \varepsilon\ \Omega - \omega$. Thus, both Fisher's exact test and the classical test are admis- ce both are in C. (Received February 11, 1958.)

the Fitting of Some Contagious Distributions, S. K. Katti and John Gurland, Iowa State College.

mber of compound and generalized distributions are compared by using such char- ics as skewness, kurtosis, and the ratio of the first two frequencies. A study has also

been made of the limiting forms of the distributions. Some of these distributions have been fitted to sampled data by estimating the parameters by various methods in order to gain some empirical knowledge of the usefulness of these distributions and the relative merits or demerits of the methods of estimation (Received February 12, 1958.)

20. Notes on the Spearman-Karber Procedures in Bioassay. (Preliminary Report) Byron Wm. Brown, Jr., University of Minnesota.

The maximum bias of the Spearman-Karber estimator of the L.D 50 over possible choices of dose levels is examined under various conditions on the distribution function, such as unimodality and symmetry. The maximum mean square error of the estimator is examined also. The results are compared with actual values for several distributions. The results are also used to make some comparisons of the Spearman-Karber estimator with some commonly used parametric methods of estimating the L.D 50 (Received February 12, 1958.)

21. Biases in Prediction by Regression for Certain Incompletely Specified Models. Harold Larsen, Iowa State College, (transmitted by H. T David).

An experimenter doesn't know whether to assume a "full" population regression model $E(y_j) = \sum_{i=1}^{k} \beta_i x_{ij}$, or a "partial" population regression model $E(y_j) = \sum_{i=1}^{m} \beta_i x_{ij}$, $k \geqq m$.

He decides the matter by the natural preliminary F-test of the hypothesis that the last $(k - m)\beta_i$'s are zero. He uses the full model for subsequent predictions if the hypothesis is rejected, and uses the partial model for subsequent predictions if the hypothesis is not rejected. Call this predictor y^*.

The full model is assumed to be true, the error terms being normally distributed with zero mean. Under these assumptions the expected value of the estimator y^* is derived. The expected value of the estimated variance of y^* is also derived if a certain sometimes-pool procedure is used. (Received February 12, 1958.)

22. Independence of Statistics and Characterization of the Multivariate Normal Distribution. S. G. Ghurye, University of Chicago and Ingram Olkin, Michigan State University, (By Title).

Some of the results proved are. If x, y are independent p-dimensional random vectors and A is a non-singular matrix such that $x + y$ and $x + Ay$ are independent, then x, y are normal. If $x_1, \cdots, x_n$ are independent random vectors, $A_1, \cdots, A_n, B_1, \cdots, B_n$ are non-singular commutative symmetric matrices such that $\sum A_i x_i$ and $\sum B_i x_i$ are independent, then the x_i are normal. If $f_1(t), \cdots, f_n(t)$ are c f 's and there exist positive numbers $\alpha_1, \cdots, \alpha_n$ such that in some neighborhood of the origin $\prod f_i^{\alpha_i}(t)$ agrees with an entire function of finite order p, where p is larger than the exponent of convergence of the zeros of the function, then p cannot exceed 2. This is applied to characterize the normal distribution by the independence of a sum of independent r.v.'s (not all of which need be identically distributed) and a polynomial of special, specified type.

Let $x_1, \cdots, x_n$ be independent p-variate normal random vectors. Let $z_j = (x_{j1}, \cdots, x_{jn})$. NASC are given for the independence of (1) $q_{ij}^{(1)} = z_i A_{ij} z_j' + z_i a_i' + z_j a_j'$ and $q_{ij}^{(2)} = z_i B_{ij} z_j' + z_i b_i' + z_j b_j'$, (2) (q_{ij}) and $\sum A_i x_i'$, and (3) $\sum A_i x_i'$ and $\sum B_i x_i'$ (Received February 1, 1958.)

23. Contributions to the Theory of Rank Order Statistics—The One-Sample Case. I. Richard Savage, University of Minnesota.

The testing that a distribution has median zero against slippage is considered using the

techniques developed earlier (these *Annals*, Vol. 27 (1956) pp. 590–615, and Vol. 28 (1957) pp. 967–977). Let $Z = (Z_1, \cdots, Z_N)$ be a random vector with $Z_i = 1(0)$ if the ith largest in absolute value in a sample of N from the density $f(x)$ is positive (negative). Then

$$P(Z = z) = N! \int_{0 \leq y_1 \leq \cdots \leq y_N \leq \infty} \cdots \int \prod_{i=1}^{N} [f^{1-z_i}(-y_i) f^{z_i}(y_i)\, dy_i]$$

Conditions are found implying $P(Z = z) > P(Z = z')$ where z is derived from z' by replacing a 0 by a 1, or interchanging a 0 and 1 in z' by moving the 1 to the left. These conditions are met by the normal and other symmetric exponential distributions. (Received February 17, 1958.)

24. An Identity of Use in Non-Linear Least Squares. M. B. WILK, Bell Telephone Laboratories.

Under rather general conditions the identity $f(x) = f(x_0) + (x - x_0)f'[(x + x_0)/2]$ is a necessary and sufficient condition that $f(x)$ be a quadratic function. The identity generalizes immediately and in the same form to p variables. A procedure due to Gauss for iterative non-linear least squares fitting of observations y_i to a function $f(x_i ; \theta)$, involves essentially the repeated linear regression of $[y_i - f(x_i ; \theta_0)]$ on $[\partial f(x_i ; \theta)/\partial\theta]_{\theta=\theta_0}$ with the regression coefficient $\hat{\delta}$ giving "improved" estimates of θ by $(\theta_0 + \hat{\delta})$. The generalization to p parameters is immediate.

This process can oscillate wildly (for example, out of computer range) and does not necessarily converge. A modification of this "Linear Gauss" procedure, based on the identity above, will approximate a "Quadratic Gauss" procedure while always solving only sets of linear equations. Advantages are a damping of the oscillations of the Linear Gauss, possible decrease in the extent of computing, and possible improvements in convergence characteristics. (Received February 17, 1958.)

25. Unbiased Regression Estimators. W. H. WILLIAMS, Iowa State College.

In sample surveys one desires unbiased estimators of population characteristics such as the mean $\bar{Y}$ of a variate y, and that these estimates be made with good precision. There are many ways of improving precision, one of which is the use of auxiliary information. In particular, this information is sometimes used in a regression estimator obtained by evaluating the line of best fit at the point $\bar{X}$. The properties of this estimator are derived from the stochastic model $y_i = A + Bx_i + e_i$ where the e_i are random errors which have expectation zero, common variance and are uncorrelated with each other. The estimator $\hat{y}_b$ of the population mean $\bar{Y}$ is then of the form $\hat{y}_b = \bar{y} + b(\bar{X} - \bar{x})$ where $\bar{y}$ and $\bar{x}$ are sample means and b is the least squares estimator of the regression slope. If the paired observations $y_i x_i$ satisfy the above linear model then $\hat{y}_b$ has expectation $\bar{Y}$. However, it is often unrealistic to assume that such a model is satisfied by the data and in such an event $\hat{y}_b$ will usually be biased. For large populations the expectation of $\hat{y}_b$ is given by $\bar{Y} - \text{Cov}(xb)$ so that $\hat{y}_b$ has a bias of $-\text{Cov}(\bar{x}b)$. $\text{Cov}(\bar{x}b)$ refers to the joint distribution of $\bar{x}$ and b in random samples of size n. An unbiased estimator of $\bar{Y}$ is obtained which has favorable efficiency. This estimator is easily generalized to the multivariate situation. (Received February 20, 1958.)

26. Maximum Likelihood Estimation from Incomplete Data for Continuous Distributions. SCOTT A. KRANE, Iowa State College.

A method is given for obtaining the maximum likelihood estimates of parameters of con-

tinuous distributions from sample data which is "incomplete" due to truncation, censoring or grouping. The method may be applied to any distribution for which the likelihood equations are soluble in the complete data case. *No special functions are required.*

The likelihood equations for incomplete samples contain two types of terms: (a) the differentials of the likelihood evaluated at observed variate values x_i, and (b) integrals of the above differentials over intervals of missing variate values. The method presented replaces the integrals in (b) by weighted sums of terms similar to (a) evaluated at variate values, z_k, "representative" of the intervals of missing values The likelihood equations for the incomplete sample are then identical with those for a complete sample of values x_i and z_k. The z_k values and weights required are functions of the parameters, so that an iterative procedure is used to obtain the estimates (Received February 26, 1958)

27. Unbiased Ratio Estimators in Stratified Sampling. JOSE NIETO DE PASCUAL, (transmitted by W. H. Williams).

The paper presents some theory of unbiased ratio estimators of the population mean $\overline{Y}$, in stratified sampling, computed from samples of k drawn from each of L strata ($k \ll L$).

Two unbiased ratio estimators and their exact variances, as well as unbiased estimates of the latter, are given. The derivations follow the lines of an unbiased ratio estimator for simple random sampling, y', introduced by Hartley and Ross (*Nature* (174), August 7, 1954, p. 270). The two estimators are (a) An unbiased "separate" ratio estimator formed by obtaining the y' estimator for each stratum, and (b) An unbiased "combined" ratio estimator computed by the y' formulae from k pairs of $\bar{y}_{st}$, $\bar{x}_{st}$, where $\bar{y}_{st}$, $\bar{x}_{st}$ are the familiar unbiased estimators of $\overline{Y}$, $\overline{X}$, computed from stratified samples of one unit drawn from each of the L strata.

These two unbiased estimators are then compared with the "combined" ratio estimator (Hansen, Hurwitz, and Gurney, *J. Amer Stat. Assn.*, Vol. 41 (1946), pp 173–189), and conditions on the population characteristics are described when the unbiased estimators are more efficient. Generalizations and the special case $k = 2$ are discussed in detail. (Received February 27, 1958.)

28. On the Laws of Cauchy and Gauss. R. G. LAHA, The Catholic University of America.

The following theorems are proved: THEOREM 1. Let x and y be two independently and identically distributed random variables having a common distribution function $F(x)$. Let the quotient $\omega = x/y$ follow the Cauchy law distributed symmetrically about the origin. Then $F(x)$ has the following general propeties: (1) it is symmetric about the origin, absolutely continuous, and has a continuous probability density function $f(x) = F'(x)$; (2) the random variable x has an unbounded range; (3) the probability density function $f(x)$ satisfies the equation $\int_0^\infty f(x)f(\omega x)x\,dx = c_0/(1 + \omega^2)$ holding for all ω, where c_0 is a constant. THEOREM 2. In addition to the conditions of Theorem 1, let $F(x)$ have finite moments of all orders. Then $F(x)$ is normal. (Received February 28, 1958.)

(Abstracts of papers presented at the Gatlinburg, Tennessee Meeting of the [illegible], April 10–12, 1958.)

29. On the Simple von Neumann Model of Dynamic Economic [illegible] as a Markov Chain. (Preliminary Report) DAVID ROSENBL[illegible], [illegible] University, (By Title).

The simple von Neumann model of dynamic economic equilibrium [illegible]

which (i) there is the same number of "goods" as of basic "productive processes" and (ii) there is a single "output" for each "productive process") is simply transformed and structurally related to two stationary Markov chains. Results are obtained for aggregation and consolidation in the simple von Neumann model and these are compared and contrasted with analogous results for macro-statistical input-output formulations. (Received February 5, 1958.)

30. Tests on a Variance-Covariance Matrix. NATHAN MANTEL, National Cancer Institute.

A class of tests on the elements of the variance-covariance matrix is proposed. The class includes as a special case Pitman's Test for equality of two correlated variances. Depending on the assumptions made the test may be one for uniformity, for equality of variances or for equality of covariances. The test may be adapted so as to provide more specific contrasts. Tests on the corresponding correlation matrix through the use of either empirical or population standardizing factors are also considered.

An interesting adaptation of the procedure is one which permits testing the interaction in a two-day classification in the absence of replication.

The testing procedures depend on the fact that when the row sums of the variance-covariance matrix are equal the mean of a set of observations is uncorrelated with any of the deviations from the mean. The test is primarily one on the significance of the multiple correlation of the mean on the set of deviations. The power efficiency of the test for specific alternatives may be increased by testing the correlation of the mean with a subtest of deviations or linear combinations of deviations. Efficiency may also be increased by shifting attention from the original variables to linear transforms. In some instances a single change in sign of some of the variables can increase efficiency. (Received February 10, 1958.)

31. An Upper Bound for the Variance of Certain Statistics. WASSILY HOEFFDING, University of North Carolina.

It is shown that if X_1, X_2, X_3 are independent and identically distributed random variables, if $0 \leqq f(X_1, X_2) = f(X_2, X_1) \leqq 1$, and $Ef(X_1, X_2) = p$, then $Ef(X_1, X_2)f(X_1, X_3) - p^2 \leqq H(p)$, where $H(p) = p^{3/2} - p^2$, $\frac{1}{2} \leqq p \leqq 1$, and $H(p) = (1-p)^{3/2} - (1-p)^2$, $0 \leqq p \leqq \frac{1}{2}$. The sign of equality holds if, with probability one, $f(X_1, X_2) = g(X_1)g(X_2)$ (for $p \geqq \frac{1}{2}$) or $f(X_1, X_2) = 1 - g(X_1)g(X_2)$ (for $p \leqq \frac{1}{2}$), where $g(X)$ takes the values 0 and 1 only. This inequality implies an upper bound for the variance of the statistic

$$U = \sum_{1 \leqq i \neq j \leqq n} f(X_i, X_j)[n(n-1)]^{-1}$$

in terms of its mean. This class of statistics includes M. G. Kendall's rank correlation coefficient t and (except for a minor difference) the Cramér–von Mises statistic ω^2. In the former case the inequality has been conjectured by Daniels and Kendall. (Received February 12, 1958.)

32. On a Test for the Equality of Several Means. K. V. RAMACHANDRAN, Demographic Training and Research Centre, India, (By Title).

Let $x_{ij}(i = 1, 2, \cdots, k; j = 1, 2, \cdots, n)$ be random samples of sizes n from k univariate normal populations with means μ_i and variance $\sigma^2 (i = 1, 2, \cdots, k)$. The hypothesis $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ against H: Not H_0 is equivalent to the union of $H_{0i}: \mu_i = \mu$ (say) against $H_i: \mu_i \neq \mu$ (where μ is unknown) for every $i = 1, 2, \cdots, k$. To test $H_{0i}: \mu_i = \mu$ against $H_i: \mu_i \neq \mu$ for any given i we have the test based on $t_i = [(\bar{x}_i - \bar{x})/S][nk/(k-1)]^{1/2}$ where $n\bar{x}_i = \sum_{j=1}^n x_{ij}$, $k\bar{x} = \sum_{i=1}^k \bar{x}_i$ and $k(n-1)S^2 = \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$. We accept H_{0i} against

H_i for any given i if $| t_i | \leqq t_{\alpha/2}$ where $t_{\alpha/2}$ is the upper $\alpha/2$ per cent point of Student's t distribution with $k(n-1)$ d f Hence we accept H_0 against $H \neq H_0$ iff (if and only if) $\max_i | t_i | \leqq t_{\alpha/2}$, i e, iff $\max_i | [(\bar{x}_i - \bar{x})/S][nk/(k-1)]^{1/2} | \leqq t_{\alpha/2}$ This two-sided version of the Nair Statistic provides an alternative test in the analysis of variance situation and gives simultaneous confidence bounds on all $\mu_i - \mu (i = 1, 2, \cdots, k)$ with a confidence coefficient $1 - \alpha$ Power properties and multivariate and other generalizations of these tests are being investigated (Received February 14, 1958)

33. On a Test for the Equality of Several Variances. K V Ramachandran, Demographic Training and Research Centre, India, (By Title).

Let $x_{ij} (i = 1, 2, \cdots, k; j = 1, 2, \cdots, n)$ be random samples of sizes n from k univariate normal populations with means μ_i and variances σ_i^2 $(i = 1, 2, \cdots, k)$ The hypothesis $H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ against H:Not H_0 is equivalent to the union of $H_{0i}: \sigma_i^2 = \sigma^2$ (say) against $H_i: \sigma_i^2 \neq \sigma^2$ (where σ^2 is unknown) for every $i = 1, 2, \cdots, k$ To test $H_{0i}: \sigma_i^2 = \sigma^2$ against $H_i: \sigma_i^2 \neq \sigma^2$ for any given i we have the test based on $F_i = S_i^2 / \sum_{l=1}^{k} S_l^2$ where $(n-1)S_i^2 = \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 l \neq i$ We accept H_{0i} against H_i for any given i if $F_1' \leqq F_i \leqq F_2'$ where $\Pr\{F_i \leqq F_1' | H_{0i}\} + \Pr\{F_i \geqq F_2' | H_{0i}\} = \alpha$ and F_i has an F distribution with $(n-1), (k-1)(n-1)$ d f Hence we accept H_0 against $H \neq H_0$ iff $F_1' \leqq F_{\min} \leqq F_{\max} \leqq F_2'$, where $F_{\min} = S_{\min}^2 / \sum_{l=1, l \neq (\min)}^{k} S_l^2$ and $F_{\max} = S_{\max}^2 / \sum_{l=1, l \neq (\max)}^{k} S_l^2$, i e, iff $g_1' \leqq S_{\min}^2 / \sum_{l=1}^{k} S_l^2 \leqq S_{\max}^2 / \sum_{l=1}^{k} S_l^2 \leqq g_2'$ This two-sided version of Cochran's statistic provides an alternative test in the homogeneity of variances situation The distribution problem, power properties and multivariate and other generalizations of these tests are being investigated (Received February 14, 1958)

34. An Optimum Property of Some Bechhofer-Type Non-Sequential Multiple-Decision Rules. Wm. Jackson Hall, University of North Carolina.

R E. Bechhofer has proposed a single-sample multiple-decision procedure for ranking means of normal populations with known variances, and, with M Sobel, a procedure for ranking variances of normal populations (*Ann Math Stat*, Vol. 25 (1954), pp. 16–39 and 273–289) We assume that the sample sizes for the populations are equal, and, in the first case, that the variances are equal Their rules guarantee a correct ranking with prescribed probability when the population parameters are sufficiently distinct (in a prescribed way) This paper proves that no other rules can accomplish this with a smaller sample size, that is, their rules are "most economical". This is not true if the sample sizes are unequal, but it is true for any analogous procedure for ranking populations according to a parameter when, for each sample, there is a numerical sufficient statistic with a monotone likelihood ratio and the parameter is a location or scale (but not range) parameter in the distribution of the statistic. These results are obtained from application of "most economical decision theory" (*Ann Math Stat.*, Vol 25 (1954), p 814) (Received February 17, 1958)

35. Second Order Rotatable Designs in Three or More Factors. R. C. Bose and Norman R. Draper, University of North Carolina.

Previous attempts to obtain second-order rotatable designs for three factors made use of the regular figures in three dimensions. A new method that has been successfully developed employs various sets of points which satisfy the conditions: $\sum x^2 = \sum y^2 = \sum z^2$, $\sum x^4 = \sum y^4 = \sum z^4$, $\sum x^2y^2 = \sum y^2z^2 = \sum z^2x^2$, all odd moments up to and including order four being zero. The basic sets may be combined in various ways to give a number of infinite classes of rotatable designs, each class dependent on a parameter. This parameter may take any value in a specified range, which depends only on the number of

points in the design. By giving specific values to the parameters in the various classes, all of the second-order designs suggested in the Institute of Statistics Mimeo Series No. 149 by D. A. Gardiner, A. H. E. Grandage, and R. J. Hader were derived as special cases. An example of such a class is as follows. The $N = 20 + n_0$ points $(\pm a, \pm a, \pm a)$; $(\pm c_1, 0, 0)$, $(0, \pm c_1, 0)$, $(0, 0, \pm c_1)$; $(\pm c_2, 0, 0)$, $(0, \pm c_2, 0)$, $(0, 0, \pm c_2)$; $(0, 0, 0)$ (n_0 times) where $c_i^2 = \{N - 8a^2 \pm [N(16a^2 - N)]^{1/2}\}/12$ $(i = 1, 2)$ form a second-order rotatable design if $.0738N \geqq a^2 \geqq .0625N$. When $c_2 = 0$, the well-known cube and octahedron design, with center points, is obtained. The method has been used additionally to construct designs for both higher-order rotatability and higher dimension (number of factors). (Received February 17, 1958.)

36. A Markov Chain Resulting From a Certain Sorting Problem. A. Bruce Clarke, University of Michigan.

Consider the following sorting problem: Objects are chosen consecutively from an infinite population consisting of r different categories in proportions $p_1, p_2, \cdots, p_r \sum p_i = 1$. The objects chosen are sorted by category and placed in r piles. Periodically one of the categories $1, 2, \cdots, r$ is selected at random with probabilities $q_1, q_2, \cdots, q_r, \sum q_i = 1$, and the pile of elements of the selected category is removed from the system. Denoting the number of elements in the ith pile immediately preceding the tth pile removal by x_{ti}, the distribution of the random vector $x_t = (x_{t1}, \cdots, x_{tr})$ is studied as $t \to \infty$. This forms a stationary Markov chain. The limiting distributions of the individual components x_{ti}, $t \to \infty$, are obtained explicitly, and a recursion formula is established which leads to the limiting distribution of x_t. One result is that the mean total number of individuals in the system at any time, $E[\sum_{i=1}^r x_{ti}]$, is minimized if the probabilities q_i are chosen proportional to $\sqrt{p_i}$. (Received February 17, 1958.)

37. Fitting the Logistic by Maximum Likelihood. J. L. Hodges, Jr., University of California, (By Title).

A method is presented by means of which the maximum likelihood estimates of the logistic response function may be quickly obtained to graphical accuracy, without the use of a computing machine or special charts. The basic idea is to replace the observed response numbers by equivalent ones for which the estimates are obvious. (Received February 19, 1958.)

38. Useful Bayes Solutions for Multiple Comparisons Problems. I. (Preliminary Report) David B. Duncan, University of North Carolina.

A Bayes solution is developed for the common t-test problem of testing the hypothesis $\theta \leqq 0$ against the alternative $\theta > 0$ given observed values of x and s where x is normally distributed with θ as mean and variance σ^2 and s^2 is an independent estimate of σ^2 distributed as $\chi_\nu^2\sigma^2/\nu$. The ultimate objective is to solve many forms of multiple comparisons problems generated by the restricted products (Lehmann, *Ann. Math. Stat.*, 1957, pp. 1–25) of problems of the given form, the Bayes solutions to be obtained as corresponding products of solutions of the form developed. The loss function assumes losses proportional to $|\theta|$, the factor for type I errors being k times that for type II errors, $k \geqq 1$. The Bayes function is a normal density with mean 0 and variance $\gamma^2\sigma^2$. These functions fit, at least to a satisfactory degree of approximation, a wide variety of problems met in practice. The solution (restricted to invariant procedures) has the critical region $x/s > t$ where t is a function of the degrees of freedom ν, loss ratio k and dispersion ratio γ^2. A brief table of t with these three arguments is presented. (Research jointly supported by the U.S. Public Health Serv-

ice and by the U.S. Air Force through the Office of Scientific Research of the Air Research and Development Command.) (Received February 20, 1958)

(Abstracts of papers presented at the Cambridge, Massachusetts Meeting of the Institute, August 25–30, 1958.)

39. Determining Bounds on Integrals with Applications to Cataloging Problems. Bernard Harris.

Assume that a random sample of size N has been drawn from a multinomial population with an unknown and perhaps countably infinite number of classes The experimenter wishes to predict $d(\alpha)$, the number of classes that will be observed in a second sample of size αN, $\alpha > 1$, (or when the sample size is increased by $(\alpha - 1)N$ additional observations); and $C(\alpha)$, the coverage of a second sample (or augmented sample), where $C(\alpha) = \sum p_i$, the sum is to be taken over those classes for which at least one representative has been observed in the sample. It is shown that $Ed(\alpha) \sim d + n_1 E\{[1 - e^{-(\alpha-1)x}]/x\}$, and $EC(\alpha) \sim 1 - (n_1/N) + (n_1/N)E\{1 - e^{-(\alpha-1)x}\}$ where d is the number of classes observed, n_1 is the number of classes occurring once in the sample, and the expectation is taken with respect to a distribution function unknown to the experimenter, but estimates of the moments are available. Hence a reasonable procedure is to compute upper and lower predictors of $d(\alpha)$ and $C(\alpha)$ by determining the suprema and infima of the above expected values subject to moment constraints.

Several results are given concerning bounds on integrals subject to moment constraints, and a method of determining the sharpest bounds is shown The explicit solutions are computed for 0, 1, 2, 3 moment constraints and applied to several examples (Received January 23, 1958)

40. Single Server Queuing Processes with a Finite Number of Sources. Gerald Harrison, The Teleregister Corporation

A service system is considered which consists of a single server and a finite number of sources. The sources are assumed to be non-interacting and to have the same negative exponential idle time distribution. The service time is assumed to have an arbitrary distribution with a finite mean. There are no defections from the waiting line, and the service time is independent of the length of the waiting line. The stationary behavior of this service system is studied. The relations between load factor, mean delay, mean service time, mean source idle time, and proportion of calls delayed are obtained The length of the waiting line at instants of termination of service is a Markov chain and its stationary distribution is thus reduced to solving a system of linear equations which, because of the form of the transition matrix, reduces to a simple iterative procedure. Under the assumption of the queuing discipline of service in the order of arrival the waiting time distribution is obtained These results are specialized to the cases of constant and negative exponential service time distributions. (Received February 13, 1958.)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of The Institute news items of interest

Personal Items

Gerald D. Berndt has taken a position as Mathematician at Headquarters, Strategic Air Command. He is Deputy Chief of Programming and Consultant in mathematics and statistics to the Commander in Chief in the Directorate of Operations.

Dr. John R. Bowman has left his post as Director of Research for Mellon Institute in Pittsburgh to take up the duties of the new Associate Dean of the Technological Institute at Northwestern University in Evanston, Illinois.

Robert J. Buehler, formerly instructor in mathematics and project associate at the Naval Research Laboratory, University of Wisconsin, has been appointed Assistant Professor in the Department of Statistics and the Statistical Laboratory at Iowa State College, Ames, Iowa.

Mavis Carroll has moved with General Foods Research Center to their new laboratories in Tarrytown, New York. She was recently named Section Head in the Product Evaluation Division, which includes responsibility for subjective testing and statistical services.

Eugene Crystal is now employed by Smith Kline & French Laboratories in Philadelphia, Pennsylvania. Mr. Crystal received an M.S. degree from Rutgers University in 1957.

Dr. Herbert A. David has resigned from the University of Melbourne to accept an appointment as Professor of Statistics at the Virginia Polytechnic Institute.

Dr. Olive Jean Dunn has resigned her position as Assistant Professor of Statistics at Iowa State College, Ames, Iowa, and has accepted a position as Assistant Professor of Biostatistics, University of California at Los Angeles.

Carl H. Fischer, Professor of Actuarial Mathematics, University of Michigan, has been by Secretary of Health, Education, and Welfare Marion B. Folson to a 12-member advisory council which is to review the long-range financial position of the Social Security System.

Dr. John E. Freund, formerly of the Department of Statistics of Virginia Polytechnic Institute, is now connected with Arizona State University. His new residence is 7532 E. Holly, Scottsdale, Arizona.

Professor Bernard Friedman, formerly of the Institute for Mathematics and Mechanics at New York University, has accepted a position as Professor of Mathematics at the University of California at Berkeley.

Paul Gunther has taken a leave of absence from the Statistical Methods Section of the General Electric Company and is now at the University of Chicago working on his Ph.D. degree.

Dr. Robert G. Hoffman has resigned from his position as Statistician for the Commission on Professional and Hospital Activities and has accepted an appointment as Statistician for the J. Hillis Miller Health Center and Assistant Research Professor of Statistics, Statistical Laboratory, University of Florida.

Donald F. Mills has completed requirements for the Ph.D. degree at the University of Washington and is now an Assistant Professor of Education at Arizona State College, Tempe.

Sidney I. Neuwirth is with Johnson & Johnson, New Brunswick, New Jersey, as Manager of Operations Research. He joined them in December, 1956, to organize and administer a company-wide O.R. program.

John W. Pratt is now Assistant Professor in the new Department of Statistics at Harvard University.

Lt. (j.g.) F. Beckley Smith, Jr., expects to be released from active duty with the U. S. Navy in February, 1958 His new address will be 264 Atlanta Drive, Pittsburgh 28, Pennsylvania.

Earl A. Thomas has accepted a position as Technical Advisory to the Manager, Ballistic Missiles Division, Burroughs Research Center, Paoli, Pennsylvania.

John Tukey is in residence at the Center for Advanced Study in the Behavioral Sciences at Stanford, California, as a Fellow. He will return to Princeton University and Bell Telephone Laboratories in September, 1958.

Pearl A. Van Natta, formerly with the Denver Research Institute, has accepted the position of biostatistician with the Child Research Council in Denver.

Robert F. White has been chosen to receive the George W. Snedecor Award in Statistics for 1958 at Iowa State College by vote of the graduate faculty in statistics. The award is given annually to the person judged to be most outstanding among students at the College working toward a Ph.D. or joint Ph.D. in statistics who are expected to graduate within a specified period of time. The award consists of a year's membership in The Institute of Mathematical Statistics together with a subscription to that professional society's journal. White is currently employed as a full-time associate on statistical problems in the Agricultural Experiment Station.

William H. Williams, formerly on a Canadian Research Council fellowship, has been appointed Assistant Professor in the Department of Statistics and the Statistical Laboratory at Iowa State College, Ames, Iowa.

Leroy F. Wolins, formerly associate director of research, Science Research Associates, Chicago, has been appointed Assistant Professor in the Department of Statistics and Psychology and the Statistical Laboratory at Iowa State College, Ames, Iowa.

G. Stanley Woodson has accepted an appointment as Biostatistician for the Commission on Professional and Hospital Activities in Ann Arbor, Michigan. The new post is supported, in part, by a grant from the W. K. Kellogg Foundation.

Summer Institute on Nonparametric Statistics

With the financial support of the National Science Foundation, the Institute of Mathematical Statistics is planning a Summer Statistical Institute (SSI) at the University of Minnesota (School of Business Administration) from June 16 to July 26. The SSI is being organized by a committee appointed by the IMS consisting of J. L. Hodges, Jr., W. Hoeffding, W. Kruskal, and I. R. Savage. The committee is planning to invite several research workers, with major and continued interests in nonparametric methods, to participate. The Plan of the SSI is to have also members who are just beginning their research work or who have not in the past concentrated in the field of nonparametric inference, for example individuals who are completing their graduate work involving research in nonparametric statistics as well as other research workers whose interests are now moving towards nonparametric statistics.

The funds allotted by the National Science Foundation permit support to some advanced graduate students who are interested in pursuing their own problems in the area of NP methods. The funds may also permit the paying of travel expenses for individuals who are supported in other ways. Also the institute will welcome a few other workers who do not need any financial support.

Summer Offerings in Statistics at Iowa State College

The Department of Statistics at Iowa State College will offer six applied courses in statistical theory and methods in its two 1958 summer sessions. These courses are planned primarily for graduate students or research workers with limited mathematical backgrounds who wish to use statistical techniques intelligently for application to other fields. In addition, a course on special topics in theoretical or applied statistics may be studied at the graduate level. Senior staff members will be available during most of the summer for consultations on research or special problems.

Students may register for either or both of the six-week summer sessions: June 17–July 23 and July 23–August 29. The complete list of statistics offerings for the first session is as follows: Stat. 401, "Statistical Methods for Research Workers" (at the level of Snedecor's *Statistical Methods*); Stat. 447, "Statistical Theory for Research Workers" (mainly theory of experimental statistics at the level of Anderson and Bancroft's *Statistical Theory in Research*; Stat. 599, "Special Topics"; and Stat. 699, "Research". In the second session will be offered Stat. 402, a continuation of 401; Stat. 448, a continuation of 447; two courses in applied methods which are more specialized, Stat. 411, "Experimental Designs for Research Workers", and Stat. 421, "Survey Designs for Research Workers"; and finally Stat. 599 and 699. Additional information may be obtained from T. A. Bancroft, department head and Director, Statistical Laboratory, Iowa State College.

Rumanian Membership

The application of Rumania for membership in the International Mathematical Union (Group II) has been approved by unanimity of the voting nations. The membership of Rumania became effective on March 1, 1958.

NAS-NRC Chairmanship

Detlev W. Bronk, president of the National Academy of Sciences, has announced the appointment of Samuel S. Wilks, professor of mathematical statistics at Princeton University, as the new Chairman of the Mathematics Division of the National Academy of Sciences-National Research Council. Dr. Wilks succeeds Paul A. Smith, professor of mathematics at Columbia University, who served as Division chairman since 1955. As Chairman, Dr. Wilks will supervise the continuing activities of the Division advisory to the Federal government in matters pertaining to mathematics. The Academy-Research Council, a private organization of distinguished scientists dedicated to the furtherance of science and its use for the general welfare, is authorized by the Federal government to act, upon request, as an official adviser in all matters of scientific and technical interest. In addition, the Mathematics Division serves as a clearing house for information of concern to mathematicians throughout the United States. It constantly strives to promote mathematical research and to improve the teaching of mathematics at all levels. As part of the Academy-Research Council, whose eight divisions embrace all the natural sciences, the Mathematics Division has a unique opportunity to develop productive interchange between mathematicians and other scientists.

New Members

The following persons have been elected to membership in The Institute

November 1, 1957, to February 3, 1958

Abernathie, Donald H., M S. (Univ. of Illinois), Research Engineer, Convair Astronautics, San Diego, California; *5335 Via Bello, San Diego 11, Calif.*

Adkins, George B., M.A. (Univ. of Missouri), Chief, Mathematical Statistics Branch, Division of Nuclear Materials Management, Atomic Energy Commission, 1901 Constitution Ave., Washington 25, D. C.; *1111 Arlington Blvd., Arlington, Virginia.*

Behnken, Donald W., M.B.A. (Columbia Univ.), Graduate Assistant, Institute of Statistics, North Carolina State College, Raleigh, North Carolina; *1631 Van Dyke Ave., Raleigh, N. C.*

Bell, Alan Edward, M S. (Stanford Univ.), Research Assistant, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California; *1024 Ramona, Palo Alto, Calif.*

Bennett, William S., M.A. (Duke Univ.), Operations Analyst, Operations Research Office, 7100 Connecticut Ave., Chevy Chase, Md.

Berman, Simeon M., B.A. (City College of New York), Lecturer, Dept. of Mathematics, The City College of New York; *1086 President St., Brooklyn 25, N. Y.*

Bhapkar, Vasant Prabhakar, M.S. (Bombay Univ.), Lecturer in Statistics, Univ. of Poona, Poona 7, India; *Dept. of Statistics, Univ. of North Carolina, Chapel Hill, North Carolina.*

Bird, Marion T., Ph.D. (Univ. of Illinois), Professor of Mathematics, San Jose State College, San Jose 14, Calif.; *45 Pala Ave., San Jose 27, Calif.*

Christensen, Inge F., M.A. (Catholic Univ.), Student, Catholic University of America, Michigan Ave., N.E., Washington 17, D. C.; *2725 29th Street, N.W., Washington 8, D. C.*

Coyne, Lolafaye, M.A. (Univ. of Kansas), Statistician, Research Dept., The Menninger Foundation, 2617 W. 6th, Topeka, Kansas.

Dembiczak, Cecilia M., M.S. (Univ. of Connecticut), Research Analyst, United Aircraft Corporation, East Hartford, Connecticut; *35 Potter Road, North Haven, Conn.*

Denton, James Q., B.S. (California Institute of Technology), Research Assistant, Department of Mathematics, University of Oregon, Eugene, Oregon.

Deshpande, J. V., M.Sc. (Univ. of Poona), Lecturer in Statistics, Govt. College of Science, Nagpur (Bombay State), India.

Dongwoo, Lee, Student, College of Liberal Arts and Sciences, Seoul National University, Seoul, Korea; #101-2 Hioza-dong, Chongno-gu, Seoul, Korea.

Doyle, Francis J., Jr., B.S. (Manhattan College), Graduate Assistant, Case Institute of Technology, Cleveland 6, Ohio; *2099 Abington Road, Cleveland 6, Ohio.*

Fisz, Marek, Doctor (Univ. of Wroclaw), Extraordinary Professor, University of Warsaw, Mathematical Institute of the Polish Academy of Sciences; *ul. J. Dabrowskiego 81/51, Warsaw, Poland*

Foster, Frederic Gordon, D.Ph. (Oxford), Reader in Statistical Computing, London School of Economics and Political Science, Houghton Street, Aldwych, London W. C. 2, England.

Kada, Jimmy Masakazu, B.S. (Univ. of California, Los Angeles), Student, School of Public Health, University of California at Los Angeles, 405 Hilgard Ave., Los Angeles 24, Calif.; *243 N. Burlington Ave., Los Angeles 26, Calif.*

Kikuchi, Susumu, B.E. (Keio-Gijuku Univ.), Assistant, The Institute of Polytechnics, Osaka City University; Minamiogimachi, Kitaku, Osaka, Japan.

Kimme, Ernest G., Ph.D. (Univ. of Minnesota), 1E237, Bell Telephone Laboratories, Murray Hill, New Jersey.

Kishen, K., Ph.D. (Univ. of Lucknow), Chief Statistician to Government, U. P., Department of Agriculture, Chhota Chhattar Manzil, Lucknow, India.

Krishnaiah, Paruchuri R., M.A. (Univ. of Minnesota), Research Assistant, Bureau of Educational Research, University of Minnesota, Minneapolis 14, Minnesota.

Laha, Radha Govinda, D.Ph. (Calcutta Univ.), Research Associate, Dept. of Mathematics, Catholic University of America, Washington 17, D. C.

Levenbach, George J., M.E.E. (Univ. of Technology, Delfd), Statistician, Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

Lewis, Charles T., B.S. (Univ. of Wyoming), Analyst, Geotechnical Corporation, 114 Grand Avenue, Laramie, Wyoming; 29 Wainwright, Laramie, Wyoming.

Li, C. C., Ph.D. (Cornell Univ.), Assistant Professor, Dept. of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pa.

Murthy, Vrudhula Krishna, B.S. (Bombay Univ.), Research Assistant, Dept. of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Oldmixon, E. Robert, Engineer, Quality Control Dept. (H-34), Surface Armament Division, Sperry Gyroscope Co., Great Neck, L. I., New York.

Popper, Juliet, Ph.D. (Stanford Univ.), Research Fellow in Psychology, Department of Psychology, Indiana University, Bloomington, Indiana.

Prabhu, Narahari Umanath, M.Sc. (Univ. of Manchester), Graduate Assistant, Dept. of Mathematics, Univ. of Western Australia, Nedlands, Western Australia; *Dept. of Statistics, Karnatak University, Dharwar, (Mysore State), India.*

Rao, Vishvember P., M.A (Univ. of Bombay), Lecturer in Statistics, College of Science, Nagpur, India

Roche, John E., Ph.D (Univ. of Texas), Associate Professor, Dept. of Business Administration, Francis Hall, Texas A & M College, College Station, Texas

Rove, Gene, B S (Drexel Inst of Tech.), Engineer, Gruen Industries, Time Hill, Cincinnati, Ohio, *3671 Meadville Drive, Sherman Oaks, California.*

Scarf, Herbert E., Ph D. (Princeton Univ.), Assistant Professor, Dept. of Statistics, Stanford University, Stanford, California.

Schmetterer, Leopold Karl, Dr. rer. nat. (University Wien), Professor and Director, Institut fur Versicherungs-mathematik und Mathematische Statistik, der Universitat Hamburg, Harvestehuderweg 10, Hamburg 13, Germany

Schumann, D. E. W., Ph.D. (Virginia Polytechnic Institute), Professor, Dept. of Statistics, University of Stellenbosch, Stellenbosch, South Africa

Searle, Shayle Robert, M A. (Univ. of New Zealand), Research Statistician, New Zealand Dairy Board, P. O. Box 866, Wellington, New Zealand, and Student, Cornell University, Ithaca, New York, *Wing Hall, Cornell Univ , Ithaca, N. Y.*

Stoker, D. J., Doctor (Univ of Amsterdam), Senior Lecturer, *Dept of Statistics, University of Pretoria, Pretoria, Union of South Africa*

Strotz, Robert H., Ph D (Univ of Chicago), Associate Professor, Dept of Economics, Northwestern University, Evanston, Illinois

Watterson, Geoffrey A., B.A (Melbourne Univ), Student, Australian National University, Box 4, G P O., Canberra, A.C T , Australia

Witting, Hermann, Dr rer. nat (Mathematisches Institut, Universitat Freiburg), Dozent Institut fur Angewandte Mathematik, Universitat Freiburg, Freiburg i Br , Germany, *Dreikonigstrasse 9, Freiburg i Br , Germany*

SMU Computing Laboratory

Southern Methodist University announces the opening of a Computing Laboratory on its campus. A new building houses the Univac Scientific 1103 Computer, the Remington Rand Service Bureau and the S M.U. Computing Laboratory offices and classrooms. The computer is operated jointly by Remington Rand as a service to industry and by S.M.U. as an academic service for research and teaching.

The S.M.U. operation is associated with the University's new Graduate Research Center. Professors and students have free use of the machine for academic research and training in computer work. Training programs are available for faculty and students. Computing projects are now underway in the fields of engineering, mathematics, psychology, law, religion, management and others. S.M.U. will make the computer arrangement involving only a nominal fee for overhead, and invites inquiries leading to such use of the machine. S.M.U. regards its laboratory as a regional university computing facility.

Doctoral Dissertations in Statistics, 1957

Listed below are doctorates conferred during the year 1957 in the United States and Canada for which the dissertations were written on topics in statistics

"Estimation of Parameters of Mixed Exponentially Distributed Failure Time Distributions from Censored Life Test Data."

Paul Dixon Minton, North Carolina State College, major in experimental statistics, "Some Distributions Related to Column Totals in Sociometric Matrices."

Robert Dean Morrison, North Carolina State College, major in experimental statistics, "Some Studies on the Estimates of the Exponents in Models Containing One and Two Exponentials."

Vasant Lakshman Mote, North Carolina State College, major in experimental statistics, "An Investigation of the Effect of Misclassification on the χ^2 Tests in the Analysis of Categorical Data."

Mangalore V. Pai, Purdue, major in statistics, "Comparisons of the Methods of Classification."

Robert N. Pendergrass, Virginia Polytechnic Institute, major in statistics, "The Rank Analysis of Triple Comparisons."

Robert Richard Read, University of California, Berkeley, major in statistics, "Contributions to the Statistical Theory of Cloud Chamber Data."

Robert H. Riffenburgh, Virginia Polytechnic Institute, major in statistics, "Linear Discriminant Analysis."

Judah I. Rosenblatt, Columbia, major in mathematical statistics, "Goodness-of-Fit Tests for Approximate Hypotheses."

Mandakini Janardan Sane, University of California, Berkeley, major in statistics, "I. Locally Unbiased Tests of Composite Hypotheses with s Contraints."

James H. Stapleton, Purdue, major in statistics, "On the Theory of Asymptotic Distributions (mod 1) and Its Extension to Abstract Spaces."

James G. C. Templeton, Princeton, major in statistics, "A Test for Detecting Single-cell Disturbances in Contingency Tables."

J. B. Tysver, University of Michigan, "Inherent errors in matrices with statistical applications."

Dirk van der Reyden, North Carolina State College, major in experimental statistics, "The Use of Orthogonal Polynomial Contrasts in the Confounding of [illegible] Experiments."

J. W. Walker, University of North Carolina, "Optimal decomposition [illegible] for estimations based on grouped data."

David Zeitlen, University of Minnesota, major in mathematics; [illegible] "Behavior of Conformal Maps under Analytic Deformation of the D[illegible]."

Alexis Zinger, University of Montreal, major in mathematical [illegible] "On the C[illegible] of the Best Amongst Three Normal Populations with Known V[illegible]."

REPORT OF THE GATLINBURG, TEN[illegible] OF THE INSTITUTE

The 1958 Eastern Regional Meeting, seventy-[illegible] of Mathematical Statistics, was held in Gatli[illegible] 1958, in conjunction with the Biometric So[illegible] Region).

The following 118 members of the Instit[illegible]

G. E. Albert, Mrs. G. E. Albert, G. N. A[illegible] Atwood, R. E. Bargman, V. P. Bhapkar, [illegible] V. J. Bofinger, R. C. Bose, G. E. P. Box, [illegible]

N. Carey, Jr., Charles Carroll, D. Chaudhuri, Victor Chew, Mary Ann Cipolloni, A. Bruce Clarke, W. H. Clatworthy, Charles W. Clunies-Ross, A. C. Cohen, W. D. Commins, W. E. Conner, William S. Connor, Dennis Cooke, Richard Cornell, Jerome Cornfield, L. C. A. Corsten, Constance Cox, Edwin L. Cox, Gertrude M. Cox, J. B. Cox, Elliot M. Cramer, Jonas M. Dalton, H. A. David, Earl Diamond, N. Draper, J. R. Duffett, David Duncan, Arthur M. Dutton, Lila Elveback, Jean Engler, A. L. Finkner, Jack Fleischer, S. M. Free, Rudolph J. Freund, Donald A. Gardiner, John J. Gart, A. Garzadela, E. Gehan, S. Geisser H. Ginsburg, W. Glenn, R. Gnanadesikian, Arnold Grandage, J. E. Grizzle, Joan M. Gurian, M. A. Guzman, R. J. Hader, W. Haenszel, W. J. Hall, Eugene K. Harris, Boyd Harshbarger, Hubert Hill, Wassily Hoeffding, R. G. Hoffman, Harold Hotelling, W. H. Horton, J. F. Hudson, David Hurst, Paul E. Irick, M. A. Kastenbaum, Therese Kelleher, George Kennedy, Allyn W. Kimball, Marcus Kjelsberg, Carl F. Kossack, Roy R. Kuebler, R. G. Laha, E. L. LeClerg, G. J. Levenbach, H. L. Lucas, Eugene Lukacs, Mary Lum, Nathan Mantel, Frank Martin, Margaret P. Martin, P. A. Miller, D. F. Morrison, George Morton, V. K. Murthy, M. D. Nefzger, George E. Nicholson, Jr., Paul S. Olmstead, D. Quade, H. F. Robinson, A. C. Rohloff, J. B. Roy, Charles F. Sarle, M. A. Schneiderman, Oliver A. Shaw, S. S. Shrikhande, Paul Somerville, D. E. South, Harold Storz, R. J. Taylor, George W. Thomson, Malcolm Turner, Ronald E. Walpole, G. S. Watson, M. B. Wilk, E. J. Williams, R. Lowell Wine.

The program of the meeting was as follows:

THURSDAY, APRIL 10, 1958

9:00–10:00 A.M.—Registration
9:00–10:00 A.M.—Biometric Society Regional Advisory Board Meeting
10:00–12:00 A.M.—Non-Linear Estimation

Chairman: R. LOWELL WINE, Hollins College

1. *Some Recent Work on Non-Linear Estimation and Design*, G. E. P. BOX, Princeton University
2. *Estimation for Linear Combinations of Exponentials*, R. G. CORNELL, United States Public Health Service
3. *Non-Linear Hypotheses*, M. B. WILK, Bell Telephone Laboratories

1:45–3:45 P.M.—Design of Experiments

Chairman: W. M. CLATWORTHY, Westinghouse Electric Corporation

1. *Use of the Direct Product of Matrices in the Analysis of Factorials*, W. S. CONNOR, National Bureau of Standards and R. C. BOSE, University of North Carolina
2. *Analysis of Variance of a Randomized Block Design with Missing Observations*, W. A. GLENN and CLYDE Y. KRAMER, Virginia Polytechnic Institute

4:00–6:00 P.M.—Contributed Papers I (I.M.S.)

Chairman: DUDLEY SOUTH, University of Florida

1. *An Upper Bound for the Variance of Certain Statistics*, WASSILY HOEFFDING, University of North Carolina
2. *A Markov Chain Resulting from a Certain Sorting Problem*, A. BRUCE CLARK, University of Michigan
3. *Second Order Rotatable Designs in Three or More Factors*, R. C. BOSE and NORMAN R. DRAPER, University of North Carolina
4. *An Optimum Property of Some Bechhofer-Type Non-Sequential Multiple-Decision Rules*, WILLIAM JACKSON HALL, University of North Carolina

5 *Useful Bayes Solutions for Multiple Comparisons Problems I. Preliminary Report*, DAVID B. DUNCAN, University of North Carolina
6 *On the Laws of Cauchy and Gauss*, R. G LAHA, Catholic University
7 *Tests on a Variance-Covariance Matrix*, NATHAN MANTEL, National Cancer Institute, and SEYMOUR GEISSER, National Institute of Mental Health
8 *On the Simple von Neumann Model of Dynamic Economic Equilibrium as a Markov Chain (Preliminary Report)*, DAVID ROSENBLATT, American University (By title)
9 *On a Test for the Equality of Several Means*, D V. RAMACHANDRAM, Demographic Training and Research Centre, Bombay, India (By title)
10. *On a Test for the Equality of Several Variances*, K. V. RAMACHANDRAM, Demographic Training and Research Centre, Bombay, India (By Title)
11 *Fitting the Logistic by Maximum Likelihood*, J L HODGES, JR, University of California, Berkeley (By Title)

8:00–9:00 P.M.—Special Invited Address (The Biometric Society)

Chairman. A W KIMBALL, Oak Ridge National Laboratory

The First Decade of the Biometric Society, C I BLISS, Connecticut Agricultural Experiment Station and Yale University

9:00 P.M.—Smoker

Host The Union Carbide Nuclear Company

FRIDAY, APRIL 11, 1958

8:30–10:30 A.M.—Special Topics I

Chairman DONALD A GARDINER, Oak Ridge National Laboratory

1. *Some Topics in the Analysis of Contingency Tables*, V. L MOTE, M. V. PAVATE and R L. ANDERSON, North Carolina State College (Presented by E J Williams, A. Grandage and V J Bofinger)
2 *Optimum Allocation for Estimation of Polynomial Regression*, E J WILLIAMS, North Carolina State College

10:45 A.M.–12:15 P.M.—Contributed Papers II (The Biometric Society)

Chairman: JACK FLEISCHER, North Carolina State College

1 *Triangle, Duo-Trio, and Difference-From-Control Tests in Taste Testing*, RALPH A. BRADLEY, Virginia Polytechnic Institute
2 *A Sequential Decision Procedure for Comparing Survival Curves*, JOHN J GART, Oak Ridge National Laboratory and Virginia Polytechnic Institute
3 *On Problems in Residual Analysis*, RUDOLF J FREUND and RICHARD W VAIL, JR, Virginia Polytechnic Institute
4 *Mixed Exponential Failure Distributions*, C W CLUNIES-ROSS, Virginia Polytechnic Institute
5. *Estimation of System Reliability from Component Reliabilities*, JAMES R DUFFETT, Virginia Polytechnic Institute

1:45–3:45 P.M.—Applications in the Physical Sciences

Chairman: CARL KOSSACK, Purdue University

Statistical Activity in the AASHO Road Test, W. N. CAREY, JR., Chief Engineer for Research, and P. E. IRICK, Chief, Data Analysis Branch, Highway Research Board, American Association of State Highway Officials Road Test

4:00–6:00 P.M.—Statistical Genetics

Chairman: H. F. ROBINSON, North Carolina State College

1. *Estimation of Sperm Frequencies in Drosophila*, M. A. KASTENBAUM, Oak Ridge National Laboratory
2. *Adaptation of High-Speed Computing Machines for Empirical Selection Studies*, F. G. MARTIN, JR., and C. C. COCKERHAM, North Carolina State College
3. *Estimation and Use of Genotype-Environmental Interaction Components of Variance in Cotton Breeding*, P. A. MILLER, J. C. WILLIAMS and H. F. ROBINSON, North Carolina State College
4. *A Synthesis of Diallel Cross Methodology*, THERESE KELLEHER, United States Department of Agriculture

8:00–10:00 P.M.—Special Invited Address (I.M.S.)

Chairman: HAROLD HOTELLING, University of North Carolina

On the Construction of Error Detecting and Error Correcting Binary Codes, R. C. BOSE, University of North Carolina

SATURDAY, APRIL 12, 1958

8:30–10:30 A.M.—Special Topics II

Chairman: BOYD HARSHBARGER, Virginia Polytechnic Institute

1. *Paired Comparisons and Tournaments*, H. A. DAVID, Virginia Polytechnic Institute
2. *Dependence in Multivariate Analysis*, R. E. BARGMANN, Virginia Polytechnic Institute.

PUBLICATIONS RECEIVED

STECK, G. P., *Upper Confidence Limits for the Failure Probability of Complex Networks*, SC-4133(TR), Sandia Corporation, Albuquerque, New Mexico, $1.25.

Boletin de la Sociedad Matematica Mexicana, Sociedad Matematica Mexicana, Tacuba 5, Mexico 1, D. F. Mexico.

Anuario Estadistico de Espana, 1957, Presidencia del Gobierno, Instituto Nacional de Estadistica, Ferraz 41, Madrid, Spain.

DUBOIS, PHILIP H., *Multivariate Correlational Analysis*, Harper and Brothers, 49 East 33D Street, New York 16, New York. pp. vii-202. $4.50.

EVES AND NEWSOM, *The Foundations and Fundamental Concepts of Mathematics*, Rinehart and Company, Inc., 232 Madison Ave., New York 16, New York. ix-363, $6.75.

DORFMAN, ROBERT, PAUL A. SAMUELSON AND ROBERT M. SOLOW, *Linear Programming and Economic Analysis*, McGraw-Hill Book Company, 330 West 42nd Street, New York 36, New York. v-527, $10.00.

System Design of Digital Computers at the National Bureau of Standards: Methods for High-Speed Addition and Multiplication. National Bureau of Standards Circular 591, issued February 14, 1958, 22 pages, 20 cents. (Order from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.)

Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues, National Bureau of Standards Applied Mathematics Series 49, issued January 15, 1958, 81 pages, 50 cents. (Order from the Superintendent of Documents, U. S. Government Printing Office, Washington, D. C.)

ASYMPTOTIC APPROXIMATIONS TO DISTRIBUTIONS[1]

By David L. Wallace

University of Chicago

1. Introduction. The study of approximations to distributions formed a major part of statistical developments during the early part of this century and included important work by Charlier, Edgeworth, Pearson and numerous others. The principal problem was the approximation to empirical distributions by theoretical functions and the methods proposed consisted chiefly either of choosing an approximating function from some class of functions, such as the Pearson type distributions or the Gram-Charlier functions, or of choosing a transformation of the variable which would reduce the distribution to approximate normality.

With the increasing importance of statistical inference, interest in the original problem of approximating to empirical distributions virtually disappeared. But interest in approximations has continued because of the increasing number and complexity of theoretical distributions and the need for usable approximations to them. In addition to the direct use for approximate evaluation of the distribution functions or the quantiles of complicated distributions, approximations have been valuable in such problems as the Behrens-Fisher problem and in the investigation of robustness of standard tests of hypotheses.

There are several general approaches to distribution approximations. The one to which I restrict attention is that of finding asymptotic expansions—in which the errors of approximation approach zero as some parameter, typically a sample size, approaches infinity. Essentially, the method consists of finding improvements to the large sample approximations used throughout statistics. A variety of expansions have been developed for many problems and the approximations are amenable to theoretical as well as empirical study.

In a simple and common form, each function $F_n(x)$ in a sequence of functions is approximated by any partial sum of a series

$$\sum_{i=0}^{\infty} \frac{A_i(x)}{(\sqrt{n})^i}$$

and the errors satisfy the condition

$$\left| F_n(x) - \sum_{i=0}^{r} \frac{A_i(x)}{(\sqrt{n})^i} \right| \leq \frac{C_r(x)}{(\sqrt{n})^{r+1}},$$

that is, the errors, using any partial sum, are of the same order of magnitude as the first neglected term. I call an asymptotic expansion valid to r terms if the

Received January 17, 1958; revised April 28, 1958.

[1] An Address presented on September 13, 1957 at the Atlantic City meeting of the Institute of Mathematical Statistics by invitation of the IMS Committee on Special Invited Papers. This paper was prepared with the partial support of the Office of Naval Research.

4:00-6:00 P.M.—Statistical Genetics

Chairman: H. F. ROBINSON, North Carolina State College

1. *Estimation of Sperm Frequencies in Drosophila*, M. A. KASTENBAUM, Oak Ridge National Laboratory
2. *Adaptation of High-Speed Computing Machines for Empirical Selection Studies*, F. G. MARTIN, JR., and C. C. COCKERHAM, North Carolina State College
3. *Estimation and Use of Genotype-Environmental Interaction Components of Variance in Cotton Breeding*, P. A. MILLER, J. C. WILLIAMS and H. F. ROBINSON, North Carolina State College
4. *A Synthesis of Diallel Cross Methodology*, THERESE KELLEHER, United States Department of Agriculture

8:00-10:00 P.M.—Special Invited Address (I.M.S.)

Chairman: HAROLD HOTELLING, University of North Carolina

On the Construction of Error Detecting and Error Correcting Binary Codes, R. C. BOSE, University of North Carolina

SATURDAY, APRIL 12, 1958

8:30-10:30 A.M.—Special Topics II

Chairman: BOYD HARSHBARGER, Virginia Polytechnic Institute

1. *Paired Comparisons and Tournaments*, H. A. DAVID, Virginia Polytechnic Institute
2. *Dependence in Multivariate Analysis*, R. E. BARGMANN, Virginia Polytechnic Institute.

PUBLICATIONS RECEIVED

STECK, G. P., *Upper Confidence Limits for the Failure Probability of Complex Networks*, SC-4133(TR), Sandia Corporation, Albuquerque, New Mexico, $1.25.

Boletin de la Sociedad Matematica Mexicana, Sociedad Matematica Mexicana, Tacuba 5, Mexico 1, D. F. Mexico.

Anuario Estadistico de Espana, 1957, Presidencia del Gobierno, Instituto Nacional de Estadistica, Ferraz 41, Madrid, Spain.

DUBOIS, PHILIP H., *Multivariate Correlational Analysis*, Harper and Brothers, 49 East 33D Street, New York 16, New York. pp. vii-202. $4.50.

EVES AND NEWSOM, *The Foundations and Fundamental Concepts of Mathematics*, Rinehart and Company, Inc., 232 Madison Ave., New York 16, New York. ix-363, $6.75.

DORFMAN, ROBERT, PAUL A. SAMUELSON AND ROBERT M. SOLOW, *Linear Programming and Economic Analysis*, McGraw-Hill Book Company, 330 West 42nd Street, New York 36, New York. v-527, $10.00.

System Design of Digital Computers at the National Bureau of Standards: Methods for High-Speed Addition and Multiplication. National Bureau of Standards Circular 591, issued February 14, 1958, 22 pages, 20 cents. (Order from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.)

Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues, National Bureau of Standards Applied Mathematics Series 49, issued January 15, 1958, 81 pages, 50 cents. (Order from the Superintendent of Documents, U. S. Government Printing Office, Washington, D. C.)

ASYMPTOTIC APPROXIMATIONS TO DISTRIBUTIONS[1]

By David L. Wallace

University of Chicago

1. Introduction. The study of approximations to distributions formed a major part of statistical developments during the early part of this century and included important work by Charlier, Edgeworth, Pearson and numerous others. The principal problem was the approximation to empirical distributions by theoretical functions and the methods proposed consisted chiefly either of choosing an approximating function from some class of functions, such as the Pearson type distributions or the Gram-Charlier functions, or of choosing a transformation of the variable which would reduce the distribution to approximate normality.

With the increasing importance of statistical inference, interest in the original problem of approximating to empirical distributions virtually disappeared. But interest in approximations has continued because of the increasing number and complexity of theoretical distributions and the need for usable approximations to them. In addition to the direct use for approximate evaluation of the distribution functions or the quantiles of complicated distributions, approximations have been valuable in such problems as the Behrens-Fisher problem and in the investigation of robustness of standard tests of hypotheses.

There are several general approaches to distribution approximations. The one to which I restrict attention is that of finding asymptotic expansions—in which the errors of approximation approach zero as some parameter, typically a sample size, approaches infinity. Essentially, the method consists of finding improvements to the large sample approximations used throughout statistics. A variety of expansions have been developed for many problems and the approximations are amenable to theoretical as well as empirical study.

In a simple and common form, each function $F_n(x)$ in a sequence of functions is approximated by any partial sum of a series

$$\sum_{i=0}^{\infty} \frac{A_i(x)}{(\sqrt{n})^i}$$

and the errors satisfy the condition

$$\left| F_n(x) - \sum_{i=0}^{r} \frac{A_i(x)}{(\sqrt{n})^i} \right| \leqq \frac{C_r(x)}{(\sqrt{n})^{r+1}},$$

that is, the errors, using any partial sum, are of the same order of magnitude as the first neglected term. I call an asymptotic expansion valid to r terms if the

Received January 17, 1958, revised April 28, 1958.

[1] An Address presented on September 13, 1957 at the Atlantic City meeting of the Institute of Mathematical Statistics by invitation of the IMS Committee on Special Invited Papers. This paper was prepared with the partial support of the Office of Naval Research.

1 partial sums have this property, and valid uniformly in x if the ·(x) do not depend on x. (The theory of asymptotic expansions is given, ıle, by Erdelyi [28].)

ıre a few points on the use of asymptotic expansions which have caused fusion. Frequently, an expansion can be extended validly to infinitely ms. For any fixed n, the infinite series may be convergent, but in applications usually is not. The asymptotic property is a property of ial sums, and though the addition of the next term will for sufficiently nprove the approximation, for any prescribed n it may not do so. the bounds $C_r(x)$ increase rapidly with r, and for small n only the first are improvements.

sharp values of $C_r(x)$ should be known. (This is rare in statistical ns but common in applications to special functions like the gamma or ıctions.) Then successive terms could be added until the error bound s minimum, giving the best guaranteed approximation, or an earlier if the error is small enough. But asymptotic expansions, except where ıt, have the inherent limitation that there is a minimum error which accuracy achievable.

aymptotic expansions used in statistics, the state of knowledge is much actory. Usually, only the order of magnitude of the errors is known, arely are explicit bounds known—and these are far from sharp. Indeed, ansions in common use have been obtained by formal operations with ected according to their order of magnitude, but without proof that the of correct order. I call these *formal* asymptotic expansions and will icate where they can be proved valid by careful but simple analysis.

proximations discussed in this paper divide into two groups, the first of approximations based ultimately on the central limit theorem and only the moments of the distribution to be approximated, and the cluding various approximations using detailed information about the ɔn.

central limit theorem. The center of a large part of the asymptotic the central limit theorem for sums of independent random variables. be a sequence of independent random variables. Denote by F_n the distribution function of the standardized sum

$$Y_n = \frac{\sum_{i=1}^{n} (X_i - E(X_i))}{\sqrt{\sum_{i=1}^{n} \mathrm{Var}\,(X_i)}}$$

the unit normal distribution function. The central limit theorem then t $\lim_{n\to\infty} F_n(x) = \Phi(x)$ for every fixed x, provided only that the means ıces are finite. If the $\{X_i\}$ are not identically distributed, an additional guaranteeing that the distributions are not too disbalanced is necessary g [51]).

st possible general results on the order of magnitude of the errors in

the central limit theorem were obtained during the 1940's by Berry [7], Esseen [29], [30], and Bergstrom [4], [5], [6]. Their results are of considerable interest and their methods are extremely important in much of asymptotic theory. The result for the sum of identically distributed random variables is that

$$\sup_x |F_n(x) - \Phi(x)| \leqq \frac{C\beta_3}{\sqrt{n}\,\sigma^3}$$

in which β_3 is the third absolute moment and σ^2 the variance of the component random variables. Several values for the constant C have been published, but only Berry's calculations have been published Hsu [45] pointed out an error in Berry's calculation. This error can be corrected without affecting the result, but there is another more serious error. I have followed through the calculation and have found that 2.05 is a satisfactory replacement for the value 1.88 given by Berry. A more careful calculation would reduce this slightly. None of the other bounds suggested is as low as 2.05. Recent work of Esseen [31] has shown that as n approaches infinity, the minimum correct value of C approaches

$$\frac{\sqrt{10} + 3}{6} \cdot \frac{1}{\sqrt{2\pi}} \approx 41$$

This value is achieved as n approaches infinity for a certain binomial distribution.

The bound holds also for sums of nonidentically distributed random variables, though the second and third moments enter in more complicated ways. Although the corrected Berry constant is the lowest known, the results of Esseen and Bergstrom are generally stronger because of the way that the second and third moments enter the bound.

All of the methods proceed by choosing as a kernel a distribution whose density function has a sharp maximum at the origin. A bound on the maximum difference of any two functions $F(x) - G(x)$ can be obtained from any bound on the convolution of this difference with the kernel distribution. The most common method of bounding the convolution has been to pass by Parseval's theorem to the characteristic functions and bound the resultant integral.

Much earlier, Lyapounov ([52], [53]) obtained a bound of order $\log n/\sqrt{n}$ for the central limit theorem error by using a normal distribution with variance of order $1/n$ for a kernel. Berry and Esseen were able to get the best result by choosing kernel distributions whose characteristic functions vanished outside a finite interval. The bounding then reduces to showing

$$\int_{-T}^{T} \frac{|f(t) - g(t)|}{|t|}\,dt = O\left(\frac{1}{T}\right)$$

where $1/T$ is the order of magnitude desired for the final result and where f and g are the characteristic functions of F and G respectively. For the central limit theorem, F is F_n, G is the normal distribution Φ and T is of order $\sqrt{n}$.

Bergstrom used the same choice as Lyapounov of a normal density for kernel, but he worked directly with the convolution integral. His method has proved

le in extensions to the multivariate central limit theorems and he has ([5], [6]) that the error there is again of order $1/n^{\frac{1}{2}}$. The characteristic n techniques have not here been used successfully.

le the central limit theorem is very useful theoretically and often in prac- is not always satisfactory. For small or moderate n, the errors of the approximation may be too large. Indeed, Berry's bound on the error is intolerable except for very large samples. Error bounds for special classes ributions—chiefly the binomial and Poisson distributions—have been by Uspensky [70] and others ([14], [33], [54], [55]).

dgeworth series for sums. To obtain improvements and to prepare for xpansions, it will be convenient to develop a class of formal expansions nes known as the Charlier differential series [11]. In this formal develop- the parameter n plays no role. The expansion is based on a distribution h need not be a normal distribution. Let ψ be its characteristic function } its cumulants. Let F be the distribution to be approximated, f its char- ic function and $\{\kappa_r\}$ its cumulants. By the definition of the cumulants, racteristic functions satisfy the formal identity

$$f(t) = \exp\left(\sum_{r=1}^{\infty}(\kappa_r - \gamma_r)\frac{(it)^r}{r!}\right)\psi(t).$$

Ψ and all its derivatives vanish at the extremes of the range of x and r all x in that range, then by integration by parts, $(it)^r\psi(t)$ is the char- ic function of $(-1)^r\Psi^{(r)}(x)$. Introducing the differential operator D to nt differentiation with respect to x, the formal identity corresponds term- any formal expansion to the formal identity

$$F(x) = \exp\left(\sum_{1}^{\infty}(\kappa_r - \gamma_r)\frac{(-D)^r}{r!}\right)\Psi(x).$$

n formally and apparently construct a distribution with prescribed cumu- r choosing Ψ and formally expanding.

nost important developing function $\Psi(x)$ is a normal distribution and with oice, the formal expansion had been given earlier by Chebyshev [13], rth [27] and Charlier [10].

yshev and Charlier proceeded by expanding and collecting terms accord- he order of the derivatives. The resulting expansion is most commonly as the Gram-Charlier A series and is identical with the formal expansion Ψ in Hermite orthogonal functions. It is a least squares expansion in ves of the normal integral Ψ with respect to a weight function which is procal of the normal density Ψ'. In this form, the expansion was de- and studied earlier by Chebyshev [12], Gram [41] and others.

A-series converges for functions F whose tails approach zero faster than Szëgo [63] or Cramér [19]). Convergence obtains for all distributions on tervals but few others of any interest. The developing normal distribution y chosen to have the same mean and variance as the given distribution F.

This choice has no effect on convergence, though it clearly has a tremendous effect on the quality of approximation by the first few terms. Altogether, the convergence properties are of little value and the importance of the Gram-Charlier series arises from its properties as an inferior form of an asymptotic expansion.

The preferable development was done by Edgeworth as an improvement to the central limit theorem. Let the distribution to be approximated again be the distribution F_n of the standardized sum Y_n (eqn. 2.1) of independent random variables Take the component random variables identically distributed with mean μ, variance σ^2, and higher cumulants $\{\sigma^r\lambda_r\ ;\ r \geqq 3\}$. Take the developing function Ψ to be the unit normal distribution function Φ. Then the cumulant differences in the formal identity (3 1) are

$$\kappa_1 - \gamma_1 = 0 = \kappa_2 - \gamma_2$$

$$\kappa_3 - \gamma_3 = \frac{\lambda_r}{n^{r/2-1}} \qquad r \geqq 3$$

The Edgeworth series is obtained by collecting terms in the formal expansion according to powers of n, thus yielding a formal asymptotic expansion of the characteristic function of the form

$$f_n(t) = \left(1 + \sum_1^\infty \frac{P_r(it)}{n^{r/2}}\right)e^{-t^2/2}$$

with P_r a polynomial of degree $3r$ with coefficients depending on the cumulants of orders 3 through $r + 2$. If powers of Φ are interpreted as derivatives, the corresponding distribution function expansion is

$$F_n(x) = \Phi(x) + \sum_1^\infty \frac{P_r(-\Phi(x))}{n^{r/2}}.$$

It is important to note that every term beyond the normal approximation can be expressed as the product of the normal density and a polynomial in x The first few terms of the expansion are:

$$F_n(x) = \Phi(x) - \frac{\lambda_3\Phi^{(3)}(x)}{6\sqrt{n}} + \frac{1}{n}\left[\frac{\lambda_4\Phi^{(4)}(x)}{24} + \frac{\lambda_3^2\Phi^{(6)}(x)}{72}\right] + \cdots$$

In 1928, Cramér [20] proved the series valid uniformly in x, but gave no explicit bounds on errors. Apart from requiring that one more cumulant exist than used in any partial sum, the proof assumes that the characteristic function h of the component random variables satisfies the condition

(3.3) $$\limsup_{|t|\to\infty} |h(t)| < 1.$$

This is satisfied if the component distribution has an absolutely continuous part. It is not satisfied for discrete distributions and the result then is generally not true.

The elementary proofs given later by Esseen [30] and Hsu [45] use the method developed for the central limit theorem bound and amount to showing that

$$\int_{-T}^{T} \frac{| f_n(t) - g_{n,k}(t) |}{| t |} dt = O\left(\frac{1}{T}\right)$$

with $T = c(n^{\frac{1}{2}})^k$ and with $g_{n,k}(t)$ the expansion of the characteristic function through terms of order $(1/n^{\frac{1}{2}})^{k-1}$ and using cumulants through order $k + 1$. Using a Maclaurin's expansion of the characteristic function, the integral up to $n^{\frac{1}{2}}$ is easily bounded by $(c_2\beta_{k+2})/T$ with the unknown distribution entering only through the absolute moment or order $k + 2$. An efficient determination of c_2 would be extremely difficult.

Using the Cramér condition (3.3) on the characteristic function, the integral from $n^{\frac{1}{2}}$ to T is easily bounded by c_3/T. But by this evaluation, the resulting bound c_3 depends on the unknown distribution through its characteristic function and this even more seriously prevents the determination of any numerically useful bounds.

Cramér [21] also proved the validity of the asymptotic expansion for sums of non-identically distributed random variables. The conditions are somewhat more restrictive. Cramér [20] showed that the termwise differentiated Edgeworth series is a valid expansion for the density function, provided the component random variables have a density function of bounded variation. Gnedenko and Kolmogorov [40] weaken this condition. They also present most of the work of Cramér and Esseen discussed here.

Esseen [30] studied the expansion problem when the Cramér condition (3.3) on the characteristic function is not satisfied. The error in using the first approximation

$$\Phi(x) - \frac{\lambda_3 \Phi^{(3)}(x)}{6\sqrt{n}}$$

is of smaller order than $1/n^{\frac{1}{2}}$ provided only that the third moment is finite and that the distribution is not a lattice distribution. If the distribution of the component random variables is lattice, i.e., takes all probability on a set of equally spaced points, a different expansion is available. The Edgeworth density function expansion is, except for a constant multiple, a valid expansion for the jumps at each possible point. The usual Edgeworth expansion for the distribution function can be modified by the addition of terms (discontinuous) so that the resultant expansion is a valid expansion, uniformly for all x. The corrections, when evaluated at the points half-way between possible values of the standardized sum, have no effect of order $1/n^{\frac{1}{2}}$, but do for all higher orders. Thus, for example, the usual Edgeworth series when applied to a binomial or Poisson distribution and evaluated only at half-integers is correct through order $1/n^{\frac{1}{2}}$ but needs a correction of order $1/n$.

Since the Gram-Charlier A series is only a rearrangement of the Edgeworth series, its asymptotic properties follow directly. Of course, many higher terms

Gram-Charlier arrangement so bad in practice is that these extra terms involve cumulants of much higher order.

Multivariate Edgeworth series can be developed in complete analogy with the univariate expansions. Other than bounds for the normal approximation error, little theoretical work has been done with the multivariate expansions. Specifically, the multivariate Edgeworth series for sums of independent random variables has *not* been shown to be a valid asymptotic expansion. This would seem the most serious gap in theoretical knowledge of asymptotic approximations

4. General Edgeworth and Cornish-Fisher series. Many sample functions have distributions asymptotically normal for increasing sample size, but not all admit asymptotic expansions beyond the normal distribution term. Expansions can be constructed for functions, such as most functions of sample moments, behaving asymptotically as sums of independent random variables. To illustrate with the simplest example, let $H(\bar{X})$ be an arbitrary function, not depending on n, of the sample mean in a sample of size n from a population with cumulants $\{\mu, \sigma^2, \sigma^r\lambda_r\}$. The distribution of

$$W_n = \frac{\sqrt{n}(H(\bar{X}) - H(\mu))}{\sigma H'(\mu)} \tag{4.1}$$

is asymptotically unit normal, provided that $H'(\mu) \neq 0$ and H is smooth enough at μ. (The assumption $H'(\mu) \neq 0$ and its equivalent for functions of several moments rule out many interesting functions for which no general theory of asymptotic expansions is known) Assume $H'(\mu) > 0$ Then the distribution function K_n of W_n is given by

$$K_n(x) = P(W_n \leqq x)$$

$$= P\left\{\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leqq \frac{\sqrt{n}\left(I\left[H(\mu) + \frac{x}{\sqrt{n}}\right] - \mu\right)}{\sigma}\right. + O(n^{-p})$$

in which I denotes the uniquely defined function inverse to H near μ and all other solutions of the inequality are easily shown to be of higher order than any power of $1/n$. If the population satisfies the Cramér condition (3.3), the standardized mean has a valid Edgeworth expansion so that

$$K_n(x) = \Phi(u) + \sum_{1}^{k-1} \frac{P_r(-\Phi(u))}{n^{r/2}} + O(n^{-k/2})$$

in which

$$u = \frac{\sqrt{n}\left(I\left[H(\mu) + \frac{x}{\sqrt{n}}\right] - \mu\right)}{\sigma}.$$

Further, each derivative $\Phi^{(i)}(u)$ can be expanded in a Taylor series in $1/n^{\frac{1}{2}}$ about $n = \infty$, evaluating the derivatives of I at $H(\mu)$ from the derivatives of H at μ. If H is smooth enough at μ, and with some natural rearrangement of terms, a

valid asymptotic expansion of the same general form as the Edgeworth series for sums is obtained. I call these series also Edgeworth series.

The construction would extend directly to the multivariate expansion of r functions of r sample moments if only the multivariate Edgeworth expansion for sums were valid. The expansion for the distribution of a single function of r moments could then be easily obtained as a marginal expansion.

I know of no literature on any of these expansions for general functions. Hsu [45] and Chung [17] proved respectively that the sample variance and the one sample t-statistic have valid expansions. (There are several errors in Chung's explicit expansion—equation (35).) Hsu proved several results needed for proofs for functions of any number of moments. But a very large amount of work was involved in completing the proof for each separate function. Hsu stated that students were working on other sample function, but I know of no others published except for a statement by Sun [62] that he had proved the result for the third moment about the mean and a proof by Hsu [46] for the expansion of the distribution of the ratios of two independent means.

A general result would be highly desirable or else an example of a statistic, smooth enough at the population value, but for which the series is not a valid asymptotic expansion to show that the construction described is not valid as generally as appears plausible.

The expansions can be obtained formally by a different approach using the Charlier differential series identity (3.2) and the classical so-called δ-method for calculating moments. Formally compute the moments and from them the cumulants of the statistic W_n of equation (4.1) by expanding $H(\bar{X})$ in a Taylor series in $\bar{X} - \mu$ and integrating term by term. The formal cumulant expansions for W_n are of the form:

$$\kappa_1(W_n) = 0 + O\left(\frac{1}{\sqrt{n}}\right)$$

$$\kappa_2(W_n) = 1 + O\left(\frac{1}{n}\right)$$

$$\kappa_r(W_n) = O(n^{-r/2+1}) \qquad r > 2$$

so that the leading terms behave exactly as for standardized sums of random variables. If these formal cumulant expansions are substituted in the symbolic identity (3.2), using the unit normal as the developing function, and if the exponential operator is expanded formally and terms collected according to powers of $n^{\frac{1}{2}}$, the same expansion as previously constructed is obtained.

This latter method is almost always easier to use in practice, especially for functions of several moments. Most applications of Edgeworth series use this method or some slight variation of it, such as using exact or valid expansions for the moments, which are frequently obtainable.

The δ-method is often used to obtain formal asymptotic expressions for moments and cumulants of statistics. A few examples of such use—for various pur-

poses—will be found in references [25], [26], [42], and [76]. The δ-method moment expansions are known to be valid in some special cases. If a function of sample moments is uniformly bounded by a power of the sample size, is smooth enough at the population moments, and if enough population moments (far more than apparently needed) exist, then the expansion can be proved valid by extending Cramér's proof [23, p. 354] for the leading terms of the mean and variance. Under severe distributional assumptions (for example, for functions of (normal theory) mean square variates—see section seven), the method can be shown valid. But there are also examples where the method is not valid, and a wide range of applications in between. However, as long as the moments are used only to get distribution approximations, it is generally plausible and sometimes known to be true that the distribution approximations are valid whether the moment expansions are or are not.

In many statistical applications, quantiles of a distribution are needed. From an Edgeworth expansion of a distribution function F_n, as asymptotic expansion for a quantile x of F_n in terms of the corresponding normal quantile z can be obtained by formal substitutions, Taylor expansions, and identification of coefficients of powers of n. The expansion is of the form

$$x = z + \frac{S_1(z)}{\sqrt{n}} + \frac{S_2(z)}{\sqrt{n}} + \cdots$$

in which the $\{S_i\}$ are polynomials. The reverse expansion

$$z = x + \frac{R_1(x)}{\sqrt{n}} + \frac{R_2(x)}{\sqrt{n}} + \cdots \tag{4.2}$$

is obtained as an intermediate step and is often useful in itself, giving an asymptotic transformation of a variate x with distribution F_n into a unit normal deviate. An expansion of the type (4.2) is often called a normalization formula. Numerically it serves the same purpose as the Edgeworth expansion but is often more convenient and possibly more accurate.

Cornish and Fisher [18] carried out these inversions, treating each cumulant of F_n according to the order of magnitude of its leading term as determined by the δ-method. For the expansion of x in terms of z, they table, for seven common probability levels, all the polynomials needed to obtain all terms through order $1/n^2$, that is, using up through the sixth cumulants.

For an absolutely continuous distribution, both of the inverted series, which I will call Cornish-Fisher series, can be proved to be valid asymptotic expansions for every probability level, whenever the initial Edgeworth series is valid. I know of no published proof of this, though Wasow's [73] proof of the invertability of a special class of distribution expansions can be modified and extended to work here.

The Edgeworth and Cornish-Fisher approximations have some faults which show up in the tails of the distribution. The distribution function approximations are not probability distributions and both monotonicity and the 0–1 range prop-

erty are violated in parts of one or both tails. Similarly the quantil mations are not always monotone in the probability levels. These trou contradict the uniform validity of the Edgeworth expansion becau refers to the absolute difference of two functions each approaching zer and not to the relative error. The validity of the Cornish-Fisher series for the probability level in each interior interval but the error increa level approaches 0 or 1.

Cramér [22], and others [15], [32], [61] in important work have in the relative accuracy of the central limit theorem approximation, b Edgeworth and Cornish-Fisher approximations, the importance of difficulties at present must be determined from empirical evidence. So ent expansions constructed to eliminate the tail difficulty will be di section six for several specific distributions.

There have been only a few numerical evaluations of the accurac approximations, largely because of the difficulty of obtaining exact comparison.

In a major piece of unpublished work, Teichroew [65] has used the the Cornish-Fisher series through n^{-5} to evaluate the quantiles of t theory chi-square distribution for a variety of degrees of freedom and p levels. He has found that the accuracy of this approximation for four freedom, provided that the probability level is not in the extreme h percent, is at least three decimals with the accuracy improving rapi degrees of freedom increase. Even for two degrees of freedom, the se curate to two decimals except in the extreme one percent. The series is the most accurate application known.

For the standardized sums of samples of size ten from four symmet normal populations, Chand [9] compared the exact quantiles with the Fisher approximations through orders $1/n$ and $1/n^2$. The latter gave be three decimal accuracy and the former better than two decimal acc probability levels ranging from $\frac{1}{2}\%$ to 25%.

Many more empirical studies of accuracy would be desirable includin of the comparative accuracies of the Edgeworth expansion and the n tion expansion (4.2).

5. Investigations of robustness. Asymptotic expansions play an i part in investigations of the effect of deviations from normality (or ot lation) on the size and power of various tests. I use the null distributi one-sample t-statistic as an example. Denote by F_n and G_n respect general and normal population distributions of the t-statistic in sa size n. Formal Edgeworth expansions of F_n and G_n can be obtained an t-statistic (but not otherwise) these have been proved valid. Since the d $F_n(x) - G_n(x)$ here is of interest, the difference of the two expansions

in which the leading term is G_n (assumed known). The approximations to F_n are then exact for a normal population and greatly improved for "near normal" populations. Similar modifications of successively higher order terms might be expected to give improved accuracy, especially for small n. The possibility (quite generally) of using expansions

$$F_n(x) \sim \sum_{i=0} B_i(x)H_i(x, n),$$

asymptotically equivalent (at every partial sum) to

$$F_n(x) \sim \sum_{i=0} A_i(x)n^{-i/2}$$

is a powerful tool to permit improved accuracy of expansions. There is no theory on how to choose good functions H_i, but useful choices can often be made on heuristic grounds or on the basis of a few computations.

In the t-statistic example, the expansion of F_n in terms of successive derivatives of the normal theory t-distribution might appear natural. Geary [39] obtained such an expansion by formally applying the Charlier differential series (equation 3.2) with G_n as the generating distribution, collecting terms according to their orders or magnitude. The result can be proved asymptotically equivalent to the Edgeworth expansion and hence valid. Geary applies the same formal method to an F-statistic (though even the formal derivation of the Charlier identity is not valid) and Bartsch [3] applies the method to various t-type statistics.

In the most substantial investigations of this kind, Gayen ([35], [36], [37], [38]) has obtained a different asymptotically equivalent expansion for the distribution of t (as well as for two-sample t, the variance ratio, and the correlation coefficient). He has given extensive tables and graphs so his expansions are far more easily used than any alternative expansions. The expansions possess also a different asymptotic property.

There seem to be no comparisons of the quality of the several approximations. Seemingly, the only feasible method for proving the validity of any of these expansions is to show equivalence to the Edgeworth series and to prove it valid (if possible). This method would never lead to useful information on accuracy since the Edgeworth series is surely much less accurate than these modified expansions.

Although Gayen's expansions are asymptotically equivalent (in n) to the Edgeworth and other series, they have an additional property, not shared by the other series mentioned, of being a formal asymptotic expansion for any fixed finite n as the population "nonnormality" approaches zero. This is made definite by assuming that the population distribution itself can be expressed by an Edgeworth expansion in some unknown parameter m (i.e., that the population values themselves are the means of m independent "elementary errors"). The Gayen expansion is a formal asymptotic expansion in powers of $1/m^{\frac{1}{2}}$ (m does not need to be known to write down the series). This approach seems conceptually more relevant to robustness problems than asymptotic expansions in the sample size.

Theoretical study of the properties of these series would be desirable, as would some comparative computations on various approximations.

6. Quantile expansions for specific distributions. The expansions that have been considered have made use of only the moments or cumulants of a distribution. Many useful asymptotic approximations have been developed from analytic expressions for the density function of the distribution to be approximated. As practically the only distributions known analytically, normal theory distributions are the object of most of these expansions. However, the normal distribution does not here play the central role that it does in the Edgeworth theory.

Consider first the expansion of a quantile of one distribution of a convergent sequence in terms of the corresponding quantile of the limiting distribution or the reverse expansion. When the normal distribution is the limiting distribution, the results are necessarily exactly those given by the Cornish-Fisher expansions but use of the explicit analytic form greatly simplifies the derivation of higher order terms and proofs of validity.

Let $\{f_n\}$ be a sequence of density functions which converges to a density function ψ. The desired expansions are found as the solutions either for t or for z of the equation

$$\int_{-\infty}^{t} f_n(x)\,dx = \int_{-\infty}^{z} \psi(x)\,dx$$

or equivalently of the differential equation

$$f_n(t)\,\frac{dt}{dz} = \psi(z). \tag{6.1}$$

In 1923, Campbell [8] obtained a formal series solution of the differential equation for the quantiles of the χ^2 distribution in terms of those of the normal distribution. He carried the series to ten terms beyond the normal approximation. Teichroew [64] has tabled these polynomial terms and used them for the computation described in section four.

Hotelling and Frankel [44] followed the same procedure to get four correction terms for the transformation of a Student's t variate into a unit normal deviate and also for the transformation of a Hotelling's T^2 variate into a chi-square variate. They proved the validity of the expansions.

Wasow [73] has given conditions on a sequence of distributions with a normal limiting distribution such that these expansions can be validly obtained by the natural formal methods, and further that each term will be a polynomial in the variate.

The accuracy of these expansions decreases as the probability level becomes more extreme. Consider the transformation of Student's t to a normal deviate. It has the form

$$z = t\left[1 + \frac{P_2(t)}{n} + \frac{P_4(t)}{n^2} + \cdots\right]$$

in which the $\{P_i\}$ are even polynomials of the indicated order. Hotelling and Frankel observed empirically that the series is of no value for t^2 greater than n. Clearly, the expansion cannot be valid for t of the order of $n^{\frac{1}{2}}$ since, from the order of the polynomials, no term would approach zero with increasing n. The usefulness of the series for small n is severely limited.

To obtain expansions useful in the tails of the distribution, Teichroew [66] has considered a limiting process in which t and z both approach infinity with n. His results are rather spectacular.

Set $t = bn^{\frac{1}{2}} + u$ with b a constant for later choice and the variable u to be kept finite. Similarly, set $z = cn^{\frac{1}{2}} + v$. The choice $c = [\log (1 + b^2)^{\frac{1}{2}}]$ is forced by examining leading terms in the differential equation (6.1) relating z and t. The equation becomes an equation relating u and v and a formal expansion of v in terms of u is easily, though tediously, obtained:

$$v = p_1(u) + \frac{p_2(u)}{\sqrt{n}} + \frac{p_3(u)}{n} + \cdots .$$

The $\{p_i(u)\}$ are polynomials of the indicated order, respectively odd and even. The dependence on b is very complicated. The whole procedure can be reversed, treating c as fixed and getting a series for u in terms of v. In actual use, with a given value of t and n, b would be chosen so that u is made small or zero thus keeping the polynomial terms small. If u is made to be zero, all odd order polynomials vanish. For 1 degree of freedom and a selection of t values corresponding to tail probability levels ranging from $\frac{1}{4}$ to 10^{-6}, choosing b so that u is zero and using the first five non-zero terms, the approximation gives the equivalent normal deviate to better than two decimal places. The ordinary series is totally worthless.

The first term is of interest. Taking $u = 0$, $b = t/n^{\frac{1}{2}}$, it is

$$z_0 = \sqrt{n \log (1 + t^2/n)}.$$

This reduces to the usual normal approximation as n approaches ∞ with t fixed. By direct analysis, Wallace [72] has shown that for all $t > 0$ and $n \geqq 1$, it satisfies the bounds

$$-\frac{.37}{\sqrt{n}} \leqq z - z_0 \leqq 0.$$

Knowing that the first term is correct to the indicated order, the entire expansion can then be shown to be a valid asymptotic expansion, uniformly for u in any finite interval. No bounds are known beyond the first term.

Teichroew has treated the χ^2 distribution in the same way with the same spectacular results. Wallace has obtained a bound as with t for the first term approximation in the upper tail.

The method is applicable to many other distributions but I know of no further applications.

7. Laplace's method and studentization. Many calculations in stati
be reduced to the evaluation of the expected value of some function of
square variate:

$$E[f(v)] = c_n \int_0^\infty f(v)v^{(n/2)-1}e^{-nv/2}\, dv.$$

The integral here is a special case of the integral

$$\int g(u)e^{-nh(u)}\, du.$$

Its asymptotic evaluation by Laplace's method is very important in th
of asymptotic expansions. If g and h are well-behaved functions, then for
the integral except in the neighborhood of the minimum of h is relatively
ble to an exponential order in n. Valid asymptotic expansions can be obt

This integral evaluation is an important part of the method of steepest
([28], p. 38) in which the path of integration, considered in the complex
chosen to pass through a minimum of h and in such a way that the absolu
of the exponential $e^{-nh(v)}$ falls off most rapidly from its maximum. The
is then expanded by Laplace's method.

The method of steepest descent (and not just the Laplace integral eva
has been used by Daniels [24] to obtain some interesting expansions th
eralize the Edgeworth expansions for sums. They have some superior pr
but make use of explicit knowledge of the moment generating function.

In the expansion of $E[f(v)]$, a simple application of the δ-method is mu
convenient than a straight application of Laplace's method (because
constant c_n in the expansion for $E[f(v)]$). Expand $f(v)$ in a Taylor series ab
population value of v (here equal to one) and integrate term by term. If
pectation exists for sufficiently large n and if f has bounded derivatives n
then the expansion obtained is valid. Since the moments of v about it
involve several powers of $1/n$, some rearrangement is needed to get an ex
of the form

$$E[f(v)] \sim \sum \frac{A_i}{n^i}.$$

But for this last step, the development would have gone as well using
mean square variate $s = v^{\frac{1}{2}}$ as argument in the Taylor expansion and integ

This expansion method and its natural extension to functions of severa
pendent mean square variates are widely applicable in statistical work
are unusually tractable for obtaining bounds on errors of approximation
am not aware of any such bounds.

One important application is to finding the distribution function
studentized statistic $Y/v^{\frac{1}{2}}$ in which Y/σ has the known distribution func
and v is an independent mean square estimate of the squared scale fac
Then

and its expansion is obtained as described. The terms in the expansion are all linear functions of the unstudentized distribution function G and its derivatives. The expansion was first obtained, in a different way, by Hartley [43]. Moriguti [56] developed the result as given here, except that he used the root mean square as argument with a consequent unnecessary complication. (His error bound (3.2) is incorrect).

Examples of the use of the expansion to get distributions of various studentized statistics are found in references [34], [58], [59], [60], and [69]. Ito [47] develops an example of a generalization to multivariate studentization.

8. The Behrens-Fisher problem. Another application is part of the development of what is to me the most interesting use of asymptotic expansions: the Welch solution for the Behrens-Fisher problem and the various extensions and analogous treatments of problems like finding confidence limits for variance components or for weighted averages when the weights must be estimated.

There have been a large number of papers attacking these problems, frequently repeating the same work ([1], [16], [57], [48], [49], [50], [67], [68], [71], [75], and others). Most of the work has consisted of formal expansions with no proofs that errors are really of their apparent order of magnitude and there has been some confusion as to what the expansions do provide. There have been a very few computations, and these very difficult, that indicate the accuracy of the approximations.

I consider in some detail a reduced form of the Behrens-Fisher problem. Let Y be normally distributed with mean μ and variance $\sum \lambda_i\sigma_i^2$ with $\{\lambda_i\}$ known positive constants and with the unknown variances σ_i^2 estimated by independent mean square variates s_i^2 respectively with n_i degrees of freedom. The problem *is to find a test of the hypothesis* $\mu = 0$, *which has significance level* α *identically* in the parameters σ_i^2.

The problem is already reduced to the sufficient statistics. Restrict it further by considering only one-sided tests of the form: reject if $Y > h(s_i^2, \alpha, \lambda_i, n_i)$ with h chosen so that $P(Y > h(s^2)) \equiv \alpha$. The Welch solution consists in an expansion

$$h(s^2) = h_0(s^2) + h_1(s^2) + h_2(s^2) + \cdots \tag{8.1}$$

in which $h_i(s^2)$ is of order n^{-i} in the degrees of freedom.

To my knowledge, it is still not known whether a non-randomized similar level α test exists. If there is no function h, the asymptotic expansion (8.1) cannot be valid. But the expansion is still of value because it provides tests that are asymptotically similar, that is, such that

$$P\left(Y > \sum_0^{r-1} h_i(s^2)\right) = \alpha + O(n^{-r}).$$

This interpretation of the asymptotic property was not clear in the original papers and was the source of confusion. From the large sample result that $Y/(\sum\lambda_i s_i^2)^{\frac{1}{2}}$ is asymptotically normally distributed the first term must be $h_0(s^2)$

$= z(\sum \lambda_i s_i^2)^{\frac{1}{2}}$ with z the level α normal quantile. Further terms are det successively so that

$$P\left(Y \leqq \sum_0^{r-1} h_i(s^2)\right) = 1 - \alpha + O(n^{-r}),$$

For getting several terms the formal operator formula given by Welch is p the most efficient procedure. The work is straightforward but Aspin [1] r that 100 pages of detailed algebra were required to determine the term h_4

I take a method that illustrates how a proof of validity could be giv determine only one term. Suppose first that $h_1(s^2)$ is any function of the va such that it and all its partial derivatives through order two are of ord Let $Q(\sigma^2) = (\sum \lambda_i \sigma_i^2)^{\frac{1}{2}}$ and let

$$u_1(s^2) = \frac{h_0(s^2) + h_1(s^2)}{Q(\sigma^2)},$$

$$P(Y \leqq h_0(s^2) + h_1(s^2)) = E\{\Phi(u_1(s^2))\}.$$

To evaluate $E\{\Phi(u_1(s^2))\}$ expand $\Phi(u_1(s^2))$ in a Taylor series in $u_1(s^2)$ – integrate with respect to the distributions of the s_i^2.

$$E\{\phi(u_1(s^2))\} = 1 - \alpha + \phi(z)E[u_1(s^2) - z] + \frac{\phi'(z)}{2} E[u_1(s^2) - z]^2$$

$$+ \frac{\phi''(z)}{6} E[u_1(s^2) - z]^3 + cE[u_1(s^2)$$

Each of these integrals here is of the form discussed in section seven validly expanded through formal Taylor expansions and termwise integ Carrying out the process far enough to get all terms of order $1/n$, and rem ing that h_1 and its derivatives are of order $1/n$, leads to the expression

$$P(Y \leqq h_0(s^2) + h_1(s^2)) = 1 - \alpha$$

$$+ \left[\phi(z) \frac{h_1(\sigma^2)}{Q(\sigma^2)} + \sum \frac{2\sigma_i^4}{n_i} \left\{ \frac{z\phi(z)}{2Q(\sigma^2)} \left[\frac{\partial^2 Q(s^2)}{\partial (s_i^2)^2}\right]_{s^2=\sigma^2} + \frac{z^2\phi'(z)}{2Q^2(\sigma^2)} \left[\frac{\partial Q(s^2)}{\partial (s_i^2)}\right]\right.\right.$$

$$+ O\left(\frac{1}{n^2}\right).$$

If h_1 is chosen to make the $1/n$ term vanish, it clearly has all the assumed erties.

The determination and proof of validity for each additional term is esse the same.

The first approximation is

$$\left[\quad (1 + z^2) \sum \frac{\lambda_i^2 s_i^4}{n_i} \right]$$

It is particularly interesting because it is equivalent, through terms of order $1/n$, to a Student's t approximation using a degrees of freedom determined by the $\{s_i^2\}$ and the $\{\lambda_i\}$ that was proposed much earlier by Welch [74].

Welch (appendix to [2]) has computed the true significance levels obtained using the expansion through h_3 for two variances, each with 6 degrees of freedom, and using a nominal level of .05. He found that the variation from .05 does not exceed .0002. This result seems quite satisfactory, but several more computations would be helpful in view of the importance of the procedure.

The theory of the expansions used by Welch and others was given in a 1949 paper by Chernoff [16]. None of the papers written on these subjects take any notice of the Chernoff work. He gives conditions for validity of expansions in an asymptotic studentization procedure due to Wald. Although his detailed results are for one nuisance parameter, he illustrates the extension to several nuisance parameters by essentially the same construction as here indicated for the Welch solution of the Behrens-Fisher problem.

A straightforward application of the Chernoff results yields an asymptotic series solution for a confidence interval for a variance component γ, where estimates v_1 of $\sigma^2 + \gamma$ and v_2 of σ^2 are available. The expansion for this problem was first developed and proved valid by Moriguti [57].

Most notable of the other work along this line is the work of James ([48], [49], [50]), who has extended the Welch formal expansion to univariate and multivariate tests of general linear hypotheses with unknown and unequal variances and covariances as nuisance parameters.

9. Conclusion. I have by no means covered all the interesting and important work on asymptotic approximations and have not even considered any non-asymptotic approaches to approximations. I have discussed what are to me some of the interesting problems, attacks, and results. Much more work is needed, particularly theoretical and empirical studies of the qualities of the approximations.

REFERENCES

[1] A. A. ASPIN, "An examination and further development of a formula arising in the problem of comparing two mean values," *Biometrika*, Vol. 35 (1948), pp 88–96

[2] A. A. ASPIN, "Tables for use in comparisons whose accuracy involves two variances, separately estimated," with appendix by B L Welch, *Biometrika*, Vol 36 (1949), pp. 290–296.

[3] G. BARTSCH, "Confidence intervals for means," unpublished doctoral dissertation Johns Hopkins University (1957).

[4] H BERGSTROM, "On the central limit theorem," *Skand. Aktuarietidskr.*, Vol. 27 (1944), pp. 139–153.

[5] H. BERGSTRÖM, "On the central limit theorem in the space R_k, $k > 1$," *Skand. Aktuarietidskr.*, Vol. 28 (1945), pp. 106–127.

[6] H. BERGSTRÖM, "On the central limit theorem in case of not equally distributed random variables," *Skand. Aktuarietidskr.*, Vol. 32 (1949), pp. 37–62.

[7] A. C. BERRY, "The accuracy of the Gaussian approximation to the sum of independent variates," *Trans. Amer. Math. Soc.*, Vol. 49 (1941), pp. 122–136.

[8] G. A. CAMPBELL, "Probability curves showing Poisson's exponential summation," *Bell System Tech. J.*, Vol. 2 (1923), pp. 95–113.
[9] U. CHAND, "Formulas for the percentage points of distributions of the arithmetic means in random samples from certain symmetrical universes," *J. Res. Nat. Bur. of Standards*, Vol. 43 (1949), pp. 79–80.
[10] C. V. L. CHARLIER, "Über das Fehlergesetz," *Ark. Mat. Astr. och Fys.*, Vol. 2 (1905–6), No. 8, 9 pp. (1905).
[11] C. V. L. CHARLIER, "Über die Darstellung willkürlicher Funktionen," *Ark. Mat. Astr. och Fys.*, Vol. 2 (1905–6), No. 20, 35 pp. (1906).
[12] P. L. CHEBYSHEV, "Sur le developpement des fonctions à une seule variable," *Bull. Acad. Imp. Sci. St. Petersbourg*, Series 3, Vol. 1 (1860), pp. 193–202.
[13] P. L. CHEBYSHEV, "Sur deux théorèmes relatifs aux probabilités," *Acta Math.*, Vol. 14 (1890), pp. 305–315.
[14] T. T. CHENG, "Normal approximation to the Poisson distribution and a proof of a conjecture of Ramanujan," *Bull. Amer. Math. Soc.*, Vol. 55 (1949), pp. 396–401.
[15] T. T. CHENG, "On asymptotic expansions connected with the sums of independent random variables," (In Chinese. English summary), *Acta Math. Sinica.*, Vol. 5 (1955), pp. 91–108.
[16] H. CHERNOFF, "Asymptotic studentization in testing of hypotheses," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 268–278.
[17] K. L. CHUNG, "The approximate distribution of Student's statistic," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 447–465.
[18] E. A. CORNISH AND R. A. FISHER, "Moments and cumulants in the specification of distributions," *Extrait de la Revue de l'Institute International de Statistique*, Vol. 4 (1937), pp. 1–14. Reprinted as paper No. 30 in R. A. Fisher, *Contributions to Mathematical Statistics*, New York, John Wiley, 1950.
[19] H. CRAMÉR, "On some classes of series used in mathematical statistics," *Proceedings of the Sixth Scandinavian Congress of Mathematicians, Copenhagen*, (1925), pp. 399–425.
[20] H. CRAMÉR, "On the composition of elementary errors," *Skand. Akturaietidskr.*, Vol. 11 (1928), pp. 13–74 and 141–180.
[21] H. CRAMÉR, "Random Variables and Probability Distributions," *Cambridge Tracts in Mathematics*, No. 36 (1937).
[22] H. CRAMÉR, "Sur un nouveau théorème-limite de la théorie des probabilités," *Actualités Sci. Ind.*, No. 736, Paris (1938).
[23] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton: Princeton University Press (1946).
[24] H. E. DANIELS, "Saddlepoint approximations in statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 631–650.
[25] F. N. DAVID, "The moments of the z and F distributions," *Biometrika*, Vol. 36 (1949), pp. 394–403.
[26] F. N. DAVID AND N. L. JOHNSON, "Statistical treatment of censored data. Part I," *Biometrika*, Vol. 41 (1954), pp. 228–240.
[27] F. Y. EDGEWORTH, "The law of error," *Cambridge Philos. Trans.*, Vol. 20 (1905), pp. 36–66 and 113–141.
[28] A. ERDELYI, *Asymptotic Expansions*, New York: Dover (1956).
[29] C. G. ESSEEN, "On the Liapounoff limit of error in theory of probability," *Ark. Mat. Astr. och Fys.*, Vol. 28A (1942), No. 9, 19 pp.
[30] C. G. ESSEEN, "Fourier analysis of distribution functions," *Acta Math.*, Vol. 77 (1945), pp. 1–125.
[31] C. G. ESSEEN, "A moment inequality with an application to the central limit theorem," *Skand. Aktuarietidskr.*, Vol. 39 (1956), pp. 160–170.

[32] W. Feller, "Generalization of a probability limit theorem of Cramér," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 361–372.
[33] W. Feller, "On the normal approximation to the binomial distribution," *Ann. Math Stat*, Vol. 16 (1945), pp. 319–329 and Vol. 21 (1950), 302.
[34] R. A Fisher, "Expansion of 'Student's' integral in powers of n^{-1}," *Metron*, Vol. 5 (1925), No. 3, pp 109–112.
[35] A K. Gayen, "The distribution of Student's t in random samples of any size from non-normal universes," *Biometrika*, Vol. 36 (1949), pp. 353–369.
[36] A. K Gayen, "The distribution of the variance ratio in random samples of any size from non-normal universes," *Biometrika*, Vol. 37 (1950), pp. 236–255
[37] A K Gayen, "Significance of difference between the means of two non-normal samples," *Biometrika*, Vol. 37 (1950), pp 399–408.
[38] A. K. Gayen, "The frequency distribution of the product moment correlation coefficient in random samples of any size from non-normal universes," *Biometrika*, Vol. 38 (1951), pp. 219–247.
[39] R C Geary, "Testing for normality," *Biometrika*, Vol 34 (1947), pp. 209–242.
[40] B. V Gnedenko and A. N Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, translated by K L. Chung, Cambridge: Addison Wesley Publishing Co (1954)
[41] J. P. Gram, "Über die Entwickelung reeler Funktionen in Reihen mittelst der Methode der kleinsten Quadrate," *J Reine Angew. Math.*, Vol 94 (1883), pp 41–73.
[42] J. B. S. Haldane, "Normalization of frequency distributions," *Biometrika*, Vol. 29 (1937), pp. 392–404
[43] H. O. Hartley, "Studentization," *Biometrika*, Vol. 33 (1944), pp. 173–180
[44] H Hotelling and L R. Frankel, "The transformation of statistics to simplify their distribution," *Ann. Math. Stat.*, Vol 9 (1938), pp. 87–96.
[45] P L. Hsu, "The approximate distribution of the mean and variance of a sample of independent variables," *Ann Math Stat.*, Vol 16 (1945), pp. 1–29.
[46] P. L. Hsu, "On the approximate distribution of ratios," *Ann Math Stat*, Vol. 16 (1945), pp. 204–210.
[47] K. Ito, "Asymptotic formulae for the distribution of Hotelling's generalized T_0^2 statistic," *Ann. Math. Stat*, Vol 27 (1956), pp 1091–1105.
[48] G. S. James, "The comparison of several groups of observations when the ratios of the population variances are unknown," *Biometrika*, Vol. 38 (1951), pp. 324–329.
[49] G. S James, "Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown," *Biometrika*, Vol. 41 (1954), pp 19–43.
[50] G. S James, "On the accuracy of weighted means and ratios," *Biometrika*, Vol 43 (1956), pp. 304–321.
[51] Y. W. Lindeberg, "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung," *Math. Z.*, Vol. 15 (1922), pp. 211–225.
[52] A. Lyapounov, "Sur une proposition de la théorie des probabilités," *Bull. Acad. Imp. Sci. St Petersbourg (Akad. Nauk SSSR)*, Ser. 5, Vol. 13 (1900), pp. 359–386.
[53] A. Lyapounov, "Nouvelle forme du théorème sur la limite de probabilites," *Mem. Acad. Imp. Sci. St. Petersbourg (Akad. Nauk SSSR), Classe Phys. Math.*, Ser. 8, Vol. 12 (1901), No. 5, 12 pp.
[54] H. Makabe, "A normal approximation to the binomial distribution," *Rep. Stat. Appl. Res. Un. Jap. Sci. Engrs*, Vol. 4 (1955), pp 47–53.
[55] H Makabe and H. Morimura, "A normal approximation to the Poisson distribution," *Rep Stat. Appl. Res. Un. Jap Sci. Engrs.*, Vol. 4 (1955), pp. 37–46.
[56] S. Moriguti, "A note on Hartley's formula of studentization," *Rep. Stat. Appl. Res. Un. Jap. Sci. Engrs.*, Vol. 2 (1953), pp 99–103.

[57] S. Moriguti, "Confidence limits for a variance component," *Rep. Stat. Appl. Res. Un. Jap. Sci. Engrs.*, Vol. 3 (1954), pp. 29–41.
[58] K. R. Nair, "The studentized form of the extreme mean square test in the analysis of variance," *Biometrika*, Vol. 35 (1948), pp. 16–31.
[59] K. R. Nair, "The distribution of the extreme deviate from the sample mean and its studentized form," *Biometrika*, Vol. 35 (1948), pp. 118–144.
[60] E. S. Pearson and H. O. Hartley, "Tables of the probability integral of the studentized range," *Biometrika*, Vol. 33 (1942), pp. 89–99.
[61] V. V. Petrov, "Extension of Cramér's limit theorem to non-identically distributed, independent quantities" (In Russian), Vestnik Leningrad Univ. (1953) No. 8, pp. 13–25.
[62] S. P. Sun, "On the successive approximation to the distribution of the third moment about the mean of independent variates," *Acad. Sinica Sci. Record*, Vol. 1 (1945), pp. 351–354.
[63] G. Szëgo, *Orthogonal Polynomials*, *American Mathematical Society Colloquium Publications*, Vol. 23 (1939).
[64] D. Teichroew, "A table of Campbell's polynomials for computation of probability points of the incomplete gamma distribution," Personal communication.
[65] D. Teichroew, "A table of millile probability points of the incomplete gamma distribution," Personal communication.
[66] D. Teichroew, "Generalized normalization polynomials for the t and incomplete gamma distributions," Personal communication.
[67] W. H. Trickett and B. L. Welch, "On the comparison of two means: further discussion of iterative methods for calculating tables," *Biometrika*, Vol. 41 (1954), pp. 361–374.
[68] W. H. Trickett, B. L. Welch, and G. S. James, "Further critical values for the two-means problem," *Biometrika*, Vol. 43 (1956), pp. 203–205.
[69] S. Ura, "On the power function of Welch's test procedure in the two sample problem," *Rep. Stat. Appl. Res. Un. Jap. Sci. Engrs.*, Vol. 4 (1955), pp. 1–13.
[70] J. V. Uspensky, *Introduction to Mathematical Probability*, New York: McGraw-Hill (1937).
[71] A. Wald, "Testing the difference between the means of two normal populations with unknown standard deviations," *Selected Papers in Statistics and Probability by Abraham Wald*, pp. 669–695, New York: McGraw-Hill (1955).
[72] D. L. Wallace, "Bounds for normal approximations to Student's and the chi-square distributions," Report SRC—80710Wc89, Statistical Research Center, University of Chicago (1958).
[73] W. Wasow, "On the asymptotic transformation of certain distributions into the normal distribution," *Proceedings of Symposia in Applied Mathematics VI, Numerical Analysis*, pp. 251–259, New York: McGraw-Hill (1956).
[74] B. L. Welch, "The significance of the difference between two means when the population variances are unequal," *Biometrika*, Vol. 29 (1937), pp. 350–362.
[75] B. L. Welch, "The generalization of 'Student's' problem when several different populations are involved," *Biometrika*, Vol. 34 (1947), pp. 28–35.
[76] E. B. Wilson and M. M. Hilferty, "The distribution of chi-square," *Proc. Nat. Acad. of Sci. U.S.A.*, Vol. 17 (1931), pp. 684–688.

A COMPARATIVE STUDY OF SEVERAL ONE-SIDED GOODNESS-OF-FIT TESTS[1]

By Douglas G. Chapman

University of Washington

0. Summary. Criteria for evaluating goodness-of-fit tests are reviewed and two additional criteria proposed. The several goodness-of-fit tests which have been proposed are studied in the light of these criteria. It is shown that it is relatively easy to evaluate the maximum and minimum power of those tests which are "partially ordered" against alternatives at a fixed "distance" from the hypothesis. A comparison is made of five tests on the basis of such minimum and maximum power functions.

1. Introduction. Let X be a real random variable with d.f. $F \varepsilon \Omega_2$ the class of continuous distribution functions (d.f.) on R. The aim of this paper is a comparative study of some of the distribution-free tests of the hypothesis

$$H_0:F = F_0$$

(where F_0 is completely specified), against the alternative

$$F < F_0 .$$

The class of distributions belonging to Ω_2 that are less than F_0 will be denoted by $\bar{\omega}$. (A distribution F is less than F_0 if $F(x) \leqq F_0(x)$ everywhere with the strict inequality holding on a set of positive F_0-measure.) Birnbaum and Scheuer [7] have called this problem that of testing goodness-of-fit against stochastically comparable alternatives. A list of a number of tests for this situation and for the case where the set of alternatives is $F \varepsilon \Omega_2$, $F \neq F_0$, as well as some of the considerations involved in designing such tests, have been given by Birnbaum [4].

If the goodness-of-fit test is merely a preliminary test to justify assumptions made for the purpose of further tests, its usefulness at the present time is debatable. As yet not enough is known of the effects of different types of deviations from assumptions on the behavior of statistical tests and estimates, nor of the effects of preliminary tests. Box and Andersen [9], however, have given examples which seem to indicate that the use of a preliminary test may leave the statistician in a less satisfactory position than if no preliminary test were made.

On the other hand the goodness-of-fit test is quite reasonable in validating a theoretical model. Moreover F, or functions of F, may enter into further developments of the whole problem so that it is desirable to have an explicit representation for it.

Received July 1, 1957; revised April 11, 1958.

[1] This research was initiated while the author held a Guggenheim Fellowship at the Readership in the Design and Analysis of Scientific Experiment at Oxford University. It was completed with the support of the O.N.R.

In many statistical applications tests are made for changes in a mean; eq
well changes in the whole distribution may be of interest. In this case one-
as well as two-sided alternatives to H_0 could be of interest to the statistici

In [4] Birnbaum noted that it is desirable to introduce a metric into the
of distributions and he suggested a number of possibilities. The choice o
metric is to a large extent a metastatistical consideration. However, the m

$$\rho(F, G) = \sup_{-\infty < x < \infty} | F(x) - G(x) |$$

or in the one-sided case

$$\rho^{-}(F, G) = \sup_{-\infty < x < \infty} (F(x) - G(x))$$

has been used extensively in probability and statistics. Furthermore these me
seem appropriate in several of the situations discussed above where a test
is reasonable. We will consider only these distance functions and more espe
the second which is appropriate to stochastically comparable alternatives.
study will be limited to those tests which have been proposed for this pro
and for which the distribution theory of the test under the null hypothe
known at least for the asymptotic case. For those tests that satisfy certain
criteria, the maximum and minimum large sample power for alternatives w
distance from the hypothesis is equal to Δ, is determined. This approach of
ing sharp upper and lower bounds for the power of a test for such alterna
was introduced by Birnbaum in [5].

The almost standard notation

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

and Z_α for the root of the equation

$$\Phi(x) = \alpha$$

will be used.

$E_0\,[f(X)]$ and $E_G\,[f(X)]$ will denote the expectation of the random func
$f(X)$ when X has the distributions F_0 and G respectively.

2. Criteria for tests of H_0. A test of H_0 of size α is a measurable func
$\varphi(X_1 \cdots X_n)$ or φ_n for short, from R_n to the interval (0, 1) such that

$$E_0(\varphi_n) \leqq \alpha.$$

Consider the alternative $G \varepsilon \bar{\omega}$. The power function $E_G(\varphi_n)$ will be den
$\beta_{\varphi_n}(G)$.

The properties of admissibility, consistency and unbiasedness for a test
well known. We refer to Birnbaum and Rubin [6] for the concepts of test

Since all tests we will consider are of structure (*d*) we may consider the problem in its canonical form, i.e., where F_0 is the uniform distribution on the interval (0, 1) and all distributions of $\bar{\omega}$ are restricted to the unit interval.

To emphasize this, it will be convenient to let u be a sure number in (0, 1) and U an r.v. uniformly distributed in (0, 1). It will also be convenient to denote by $U_1, U_2, \cdots, U_n$ the *ordered* sample from this distribution. In some instances it will be convenient to introduce U_0 and U_{n+1}. These are set equal to 0 and 1 respectively.

We also introduce two more concepts, monotonicity and partial ordering, as applied to tests of the hypothesis H_0.

DEFINITION 1. φ is a monotone test of H_0 if

$$(1)\quad U_i \geqq V_i \quad (i = 1, 2, \cdots n) \Rightarrow (U_1, U_2, \cdots U_n) \geqq (V_1, V_2, \cdots V_n).$$

DEFINITION 2. φ is a partially ordered [p.o.] test of H_0 if

$$(2)\qquad G_1(u) \leqq G_2(u) \qquad \text{for all} \qquad u \,\varepsilon\, (0, 1) \Rightarrow \beta_\varphi(G_1) \geqq B_\varphi(G_2).$$

From the continuity theorem for Lebesgue-Stieltjes integrals we have the following obvious

REMARK. If φ is continuous except for a finite number of jumps and φ is p.o then φ is unbiased.

The relationship between monotonicity and partial ordering will be useful later.

THEOREM. *Tests of structure (d) that are monotone are* p.o.

PROOF. Let $G_1(u) \leqq G_2(u) \leqq u$ and recall

$$(3)\qquad \beta_\varphi(G_i) = \int_0^1 \int_0^1 \cdots \int_0^1 \varphi(u_1, u_2, \cdots, u_n) \prod_{j=1}^n dG_i(u_j) \qquad (i = 1, 2).$$

Make the change of variables

$$y_j = G_i(u_j) \qquad (j = 1, 2, \cdots n)(i = 1, 2)$$

in the two integrals. The inverse is defined in the usual fashion, i.e.,

$$u_j = G_i^{-1}(y_j) = \inf_{0 \leqq x \leqq 1} [x : G_i(x) = y_j].$$

The two integrals become

$$(4)\qquad \int_0^1 \int_0^1 \cdots \int_0^1 \varphi[G_i^{-1}(y_1), \cdots, G_i^{-1}(y_n)] \prod_{j=1}^n dy_j \qquad (i = 1, 2).$$

Since $G_1 \leqq G_2$, $G_1^{-1} \geqq G_2^{-1}$; this, together with the monotonicity property of φ, implies the required inequality for $\beta_\varphi(G_1)$, $\beta_\varphi(G_2)$.

It may also be noted that any monotone test is admissible. This follows from a result of A. Birnbaum [2] (appendix) who considered this problem where the set of alternatives is restricted to d.f. with monotone densities. In this paper we de-

termine which of the several tests of H_0 that have been suggested satisfy these criteria and then determine

$$\underline{\beta}_\varphi(\Delta) = \inf_{G \varepsilon \mathcal{G}(\Delta)} \beta_\varphi(G); \qquad \bar{\beta}_\varphi(\Delta) = \sup_{G \varepsilon \mathcal{G}(\Delta)} \beta_\varphi(G),$$

where

$$\mathcal{G}(\Delta) = [G: G \,\varepsilon\, \bar{\omega}, \qquad \rho^-(F_0, G) = \Delta], \qquad\qquad 0 < \Delta < 1,$$

for these several test functions φ.

To obtain sharp upper and lower bounds of the power of any p.o. test against all alternatives G such that

$$\rho^-(F_0, G) = \Delta, \tag{5}$$

we consider the alternatives

$$G_{mu_0}(u) = \begin{cases} 0, & u < 0, \\ u, & 0 \leqq u < u_0, \\ u_0, & u_0 \leqq u < u_0 + \Delta, \\ u, & u_0 + \Delta \leqq u < 1, \\ 1, & 1 \leqq u, \end{cases} \tag{6}$$

and

$$G_M(u) = \begin{cases} 0, & u < \Delta, \\ u - \Delta, & \Delta \leqq u < 1, \\ 1, & 1 \leqq u. \end{cases} \tag{7}$$

These distributions are not members of the family of alternatives $\bar{\omega}$, but it is possible to find distributions in $\bar{\omega}$ arbitrarily close to G_{mu_0} or G_M. Hence it follows from the continuity of the power functions, that if the test is p.o.

$$\underline{\beta}_\varphi(\Delta) = \inf_{0 \leqq u_0 \leqq 1-\Delta} \beta_\varphi(G_{mu_0}); \qquad \bar{\beta}_\varphi(\Delta) = \beta_\varphi(G_M).$$

Such bounds are given below for several of the tests of H that meet the criteria of admissibility, consistency, unbiasedness, monotonicity and partial orderedness.

3. Fisher and Pearson tests. The statistics

$$\pi = -2 \sum_{i=1}^{n} \ln U_i, \tag{8}$$

$$\pi' = -2 \sum_{i=1}^{n} \ln (1 - U_i) \tag{9}$$

were introduced in the problem of combining tests but are also suitable for testing H_0. If H_0 is true π and π' both are distributed as χ^2 with $2n$ d.f. Furthermore, the u.m.p. test of H_0 against the family of alternatives

$$G_k = u^k \qquad\qquad k > 1 \tag{10}$$

is obviously of the form: Reject H_0 if $\pi < c$. A similar statement may be made about π'.

Furthermore, such tests are obviously monotone and hence p.o.

It will be convenient to refer to the tests, reject H if $\pi < c$ or $\pi' > c$, simply as the tests π, π'. These are two of the class of likelihood ratio tests of the form: Reject H if $\sum_{i=1}^{n} \ln g_1(U_i) > c$, where g_1 is the derivative of a specified absolutely continuous alternative G_1.

If $E_0[\ln g_1(U)]^2 < \infty$, this test statistic is asymptotically normal and furthermore if $E_G[\ln g_1(U)]^2 < \infty$ and $E_0[\ln g_1(U)] < E_G[\ln g_1(U)]$ the usual argument shows that the test based on $\sum_{i=1}^{n} \ln g_1(U_i)$ is consistent for testing H_0 against the alternative G.

In particular for the tests π, π', we have

THEOREM. *The tests π, π' are consistent for the set of alternatives $\bar{\omega}$.*

PROOF. In view of the remark above it is necessary to show that $E_0[\ln U]^2$, $E_G[\ln U]$ are finite and $E_0[\ln U] < E_G[\ln U]$. Now

$$E_G[\ln U]^2 = \int_0^1 (\ln u)^2 \, dG(u). \tag{11}$$

Let $1 > \epsilon > 0$; for every ϵ

$$\int_\epsilon^1 (\ln u)^2 \, dG(u) = (\ln u)^2 G(u) |_\epsilon^1 + 2 \int_\epsilon^1 G(u) \left| \frac{\ln u}{u} \right| du. \tag{12}$$

Since $G(u) \leqq u$ the first term on the right-hand side of (12) can be made arbitrarily small by appropriate choice of ϵ while for all ϵ the second integral is bounded by

$$\int_0^1 |\ln u| \, du = 1.$$

This shows that both $E_0[\ln U]^2$, $E_G[\ln U]^2$ exist and also validates the integration by parts in the next step.

For

$$\begin{aligned} E_G(\ln U) &= \int_0^1 (\ln u) \, dG(u) \\ &= \ln uG(u)|_0^1 - \int_0^1 \frac{G(u)}{u} \, du. \end{aligned} \tag{13}$$

The first term on the right-hand side of (13) is zero. Since $G(u) \leqq u$ with inequality holding on a set of positive measure

$$-\int_0^1 \frac{G(u)}{u} \, du > -\int_0^1 du = -1 = E_0[\ln U]$$

as required for the consistency of the π test.

The proof of the consistency of π' requires consideration of two cases.

CASE 1. $\qquad E_G[\ln(1 - U)] > -\infty.$

Since G is continuous and $G(u) < u$ on a set of positive measure, $\exists\ \epsilon'$ such that

$$\int_0^{1-\epsilon'} \frac{[u - G(u)]}{1 - u}\, du = 2\delta > 0.$$

Now in view of the finiteness of $E_G[\ln (1 - U)]$, $\exists\ \epsilon$ which may be chosen less than $\delta/3$ and ϵ' such that

$$|\ln \epsilon|\, [1 - G(1 - \epsilon)] < \delta/3$$

and also

$$\left| \int_0^1 \ln (1 - u)\, dG(u) - \int_0^{1-\epsilon} \ln (1 - u)\, dG(u) \right| < \delta/3.$$

Now

$$\begin{aligned}\int_0^1 \ln (1 - u)\, dG(u) &< \int_0^{1-\epsilon} \ln (1 - u)\, dG(u) + \delta/3 \\ &= G(1 - \epsilon) \ln \epsilon + \int_0^{1-\epsilon} \frac{G(u)}{1 - u}\, du + \delta/3 \\ &= G(1 - \epsilon) \ln \epsilon + \int_0^{1-\epsilon} \frac{u}{1 - u}\, du - \int_0^{1-\epsilon} \frac{u - G(u)}{1 - u}\, du + \delta/3 \\ &= -1 + \epsilon + \ln \epsilon[G(1 - \epsilon) - 1] - 2\delta + \delta/3 < -1 - \delta.\end{aligned}$$

Since the critical region of the π' test converges to: Reject H_0 if

$$-\frac{\pi'}{2n} < -1 - \frac{Z_\alpha}{\sqrt{n}},$$

while by Khintchine's theorem $-(\pi'/2n)$ converges almost surely to $E_G[\ln (1 - U)] \leqq -1 - \delta$ under the alternative G, the consistency follows.

CASE 2. $E_G[\ln (1 - U)] = -\infty$.

By well-known results in this case infinitely many of the sequence of the independent r.v. $\sum_{i=1}^{n} \ln (1 - U_i)\ n = 1, 2, 3 \cdots$, are with probability 1 less than nA for any arbitrary A. Hence from the remark on the critical region the consistency is immediate.

As a consequence of this theorem it may be noted that π is asymptotically normal both under H_0 and all alternatives in $\bar{\omega}$; it is trivial to give examples that this is not true for π'. This behavior is reversed for alternatives $G(u) \geqq u$.

The asymptotic normality of π permits an elementary derivation of $\beta_\pi(\Delta)$ and $\bar{\beta}_\pi(\Delta)$ for large samples. In particular

(14) $$E_M |\ln U| = 1 - \Delta(1 - \ln \Delta),$$

(15) $$\sigma_M^2 |\ln U| = 1 + 2\Delta^2 \ln \Delta - (\ln^2 \Delta)(\Delta + \Delta^2),$$

and

(16) $$E_{mu_0}|\ln U| = 1 - \Delta + u_0 \ln\left(1 + \frac{\Delta}{u_0}\right)$$

(17) $$\text{with} \quad \max_{u_0} E_{mu_0}|\ln U| = (1 - \Delta)[1 - \ln(1 - \Delta)].$$

This maximum is attained when $u_0 = 1 - \Delta$. Also

(18) $$E_{Mu_0}(\ln U)^2 = 2(1 - \Delta) - u_0[\ln^2(u_0 + \Delta) - \ln^2 u_0] - 2(u_0 - \Delta)\ln(u_0 + \Delta) - 2u_0 \ln u_0 .$$

A numerical study of the variance of $\ln U$ as a function of u_0 shows that the variance is maximized when $u_0 = 1 - \Delta$ though the changes with respect to u_0 are very slight.

For $u_0 = 1 - \Delta$

(19) $$\sigma_m^2(\ln U) = (1 - \Delta)[2 - 2\ln(1 - \Delta) + \ln^2(1 - \Delta)] - (1 - \Delta)^2[1 - \ln(1 - \Delta)]^2.$$

Hence approximately for large n

(20) $$\beta_\pi(\Delta) = \Phi\left(\frac{[Z_\alpha + \sqrt{n}\,[(1 - \Delta)\ln(1 - \Delta) + \Delta]}{\sigma_m(\ln u)}\right),$$

(21) $$\bar{\beta}_\pi(\Delta) = \Phi\left(\frac{Z_\alpha + \sqrt{n}\,[\Delta(1 - \ln\Delta)]}{[1 + 2\Delta^2 \ln\Delta - \ln^2\Delta(\Delta + \Delta^2)]^{1/2}}\right).$$

The minimum power of the π' test is attained against the alternative G_{m0}, i.e., the jump of height Δ is located at $u = \Delta$. Furthermore this minimum power is the same as the minimum power of the π test.

On the other hand π' will not be asymptotically normally distributed for G_M; in fact with probability $1 - (1 - \Delta)^n$, $\pi' = +\infty$ in which case rejection is immediate. However, under the condition that all the U_i are less than 1, π' is asymptotically normal so that

(22) $$\bar{\beta}_{\pi'}(\Delta) \doteq 1 - (1 - \Delta)^n[1 - \Phi(x)],$$

where

(23) $$x = \frac{Z_\alpha - \sqrt{n}\,\dfrac{\Delta \ln \Delta}{1 - \Delta}}{\left[1 - \dfrac{\Delta \ln^2 \Delta}{(1 - \Delta)^2}\right]^{1/2}}.$$

Tables giving numerical values of these minimum and maximum power functions are displayed in section 8 below where the several tests are compared.

3. D_n^- **test.** The empirical d.f. $F_n(u)$ is basic in many distribution-free tests of H_0. The use of the statistic

(24) $$D_n^- = \sup_{0 \leq u \leq 1} [u - F_n(u)]$$

as a large sample test for H_0 became possible after Smirnov [20] obtained its limiting distribution. Subsequently Birnbaum and Tingey [3] gave a closed expression for the distribution of D_n^- for finite n.

These results are

$$\lim_{n\to\infty} \Pr[\sqrt{n}\, D_n^- \leqq Z] = 1 - e^{-2Z^2} \tag{25}$$

and

$$\Pr[D_n^- \leqq \epsilon] = 1 - \epsilon\left(\sum_{j=0}^{[n(1-\epsilon)]} \binom{n}{j}\left(1 - \epsilon - \frac{j}{n}\right)^{n-j}\left(\epsilon + \frac{j}{n}\right)^{j-1}\right), \tag{26}$$

where as usual $[x]$ is the greatest integer contained in x.

It is immediate from the definition that the test is monotone and hence p.o. and admissible, as well as being consistent.

Birnbaum in [5] gave upper and lower bounds for the power of the D_n^- test for alternatives of fixed distance Δ within the class of all continuous distribution functions. The upper bound is attained for the alternative labeled here G_M and we quote his result

$$\bar{\beta}_{D_n^-}(\Delta)\begin{cases} = (\epsilon_n - \Delta) \displaystyle\sum_{i=0}^{[n(1-\epsilon_n+\Delta)]} \binom{n}{i}\left(1 - \epsilon_n + \Delta - \frac{i}{n}\right)^{n-1}\left(\epsilon_n + \Delta + \frac{i}{n}\right)^{i-1} \\ = 1 \begin{cases}\text{for } \epsilon_n \geqq \Delta, \\ \text{for } \epsilon_n < \Delta,\end{cases}\end{cases} \tag{27}$$

where ϵ_n is chosen so that

$$\Pr[D_n^- > \epsilon_n \mid H_0] = \alpha.$$

In view of Smirnov's result for large n

$$\bar{\beta}_{D_n^-} - (\Delta) \doteq e^{-2n(\epsilon_n-\Delta)^2} \qquad \text{for } \epsilon_n \geqq \Delta.$$

The lower bound of the power of the D_n^- test within the class of stochastically comparable alternatives was studied by Birnbaum and Scheuer [7]. Their result is given as a number of double and triple sums of terms of the same type as those in (26), and is not in a form useful for comparison or evaluation purposes.

The following approach does not yield a simple closed expression for the exact power, but an adequate approximation is obtained. We write

$$\begin{aligned}\beta(G_{mu_0}) &= \Pr[u_0 + \Delta - F_n(u_0 + \Delta - 0) \geqq \epsilon_n \mid G_{mu_0}) \\ &+ \Pr[\sup_{0\leqq u<u_0} \{u - F_n(u)\} \geqq \epsilon_n \mid u_0 + \Delta - F_n(u_0 + \Delta - 0) < \epsilon_n, G_{mu_0}] \\ &+ \Pr[\sup_{u_0+\Delta\leqq u<1} \{u - F_n(u)\} \geqq \epsilon_n \mid \sup_{0\leqq u_0<u_0+\Delta} \{u - F_n(u)\} < \epsilon_n, G_{mu_0}].\end{aligned} \tag{28}$$

It will be convenient to symbolize the three terms on the right hand of (28) by P_1, P_2, P_3 respectively. It is immediate that

$$P_1 = \sum_{k=0}^{[n(u_0+\Delta-\epsilon_n)]} B(k; n, u_0), \tag{29}$$

where the right-hand summands denote binomial probabilities in the usual notation.

An examination of the integral representation of

$$P_n(\epsilon) = \Pr\left[\sup_{0 \leqq u \leqq 1} \{u - F_n(u)\} \leqq \epsilon\right]$$

given by Birnbaum and Tingey in [3]

$$P_n(\epsilon) = n! \int_0^{\epsilon} \int_{x_1}^{(1/n)+\epsilon} \int_{x_2}^{(2/n)+\epsilon} \cdots \int_{x_K}^{(K/n)+\epsilon} \int_{x_{K+1}}^{1} \cdots \int_{x_{n-1}}^{1} dx_n \cdots dx_{K+2}\, dx_{K+1} \cdots dx_3\, dx_2\, dx_1,$$

where $K = [n(1 - \epsilon)]$, shows immediately that P_2 and P_3 are bounded by α. Hence the dominant term in $\beta(G_{mu_0})$ is P_1 which is minimized when $U_0 = \frac{1}{2}$. This value has been used in making minimum power calculations for the D_n^- test

However the actual values of P_2 can be determined in the large sample case. Consider

$$\Pr\left[\sup_{0 \leqq u \leqq u_0} \{u - F_n(u)\} \leqq \epsilon_n \,\middle|\, F_n(u_0) = \frac{k}{n}, G_{mu_0}\right]$$

$$= k! \int_0^{\epsilon_n} \int_{u_1}^{1/u_0[(1/n)+\epsilon_n]} \cdots \int_{u_{K'}}^{1/u_0[(K'/n)+\epsilon_n]} \int_{u_{K'+1}}^{1} \cdots \int_{u_{k-1}}^{1} du_k \cdots du_{K'+2}\, du_{K'+1} \cdots du_2\, du_1, \tag{30}$$

where $K' = [n(u_0 - \epsilon_n)]$.

The integral form can be written down in a similar manner to that of $P_n(\epsilon)$ and the result given is obtained by a trivial change of variable. By a slight extension of the arguments used by Birnbaum and Tingey this can be expressed as a closed sum, viz

$$1 - \left(\frac{n\epsilon_n}{k}\right)\left(\frac{k}{nu_0}\right)^k \sum_{j=0}^{K'} \binom{k}{j} \left(\frac{nu_0}{k} - \frac{n\epsilon_n}{k} - \frac{j}{k}\right)^{k-j} \left(\frac{n\epsilon_n}{k} + \frac{j}{k}\right)^{j-1}. \tag{31}$$

It is convenient to denote the function on the right-hand side

$$V(nu_0/k,\, n\epsilon_n/k,\, k).$$

The power of the D_n^- test against alternatives of the form

$$G_a(u) = \begin{cases} au, & 0 \leqq u < 1, 0 < a < \infty, \\ 1, & u \geqq 1, \end{cases}$$

can be expressed in terms of the function V. This can be seen by writing down the integral using the general power formula given by Birnbaum ([5], p. 486) or by a simple direct argument. In fact

$$\beta_{D_n^-}(G_a) = 1 - V\left(\frac{1}{a}, \epsilon_n, n\right). \tag{32}$$

While the sums in (31) can be evaluated by a straight-forward process, the process is tedious for large k, and we obtain instead an asymptotic result that yields a method of approximating V in this situation.

Let $G_{na}(u)$ denote a sequence of d.f. of the form

$$(33) \qquad G_{na}(u) = \begin{cases} \left(1 + \frac{a}{\sqrt{n}}\right)^{-1} u, & 0 \leqq u < b, -\sqrt{n} < a < \sqrt{n}, \\ 1, & u \geqq b, \end{cases}$$

where

$$b = \min\left(1 + \frac{a}{\sqrt{n}}, 1\right).$$

Then

$$(34) \qquad \beta_{D_n^-}(G_{na}) = \Pr\left[\sup_{0 \leqq u < b} \left(1 + \frac{a}{\sqrt{n}}\right) u - F_n(u) > \epsilon_n\right].$$

Now we use Donsker's theorem [11] justifying Doob's heuristic approach to the Kolmogorov-Smirnov theorems [12] to validate the following steps:

$$(35) \qquad \begin{aligned} \lim_{n \to \infty} \Pr [\sup_{0 \leqq u < b} \{\sqrt{n}(u - F_n(u) + au)\} > Z] \\ = \Pr [\sup_{0 \leqq u < 1} (X(u) + au) > Z], \end{aligned}$$

where $X(u)$ is a Gaussian process with the properties noted by Doob ([12], p. 397). Further, the transformation he made and his evaluation of

$$\Pr \{\sup [\zeta(t) - (at + b)] \geqq 0\}$$

may be used to evaluate this last probability. We have in fact

$$(36) \qquad \begin{aligned} \Pr [\sup_{0 \leqq u < 1} [X(u) + au] > Z] &= \Pr\left[\sup_{0 \leqq u < \infty} \frac{\zeta(u) + au}{u + 1} > Z\right] \\ &= e^{-2Z(Z-a)}. \end{aligned}$$

In other words, putting $\epsilon_n = Z/\sqrt{n}$

$$\lim_{n \to \infty} V\left(1 + \frac{a}{\sqrt{n}}, \frac{Z}{\sqrt{n}}, n\right) = 1 - e^{-2Z(Z-a)}.$$

Hence if $n, k \to \infty$ with $\sqrt{k}[(n/2k) - 1] = (n - 2k)/2\sqrt{k}$ and n/k remaining finite and if u_0 is set equal to $\frac{1}{2}$

$$(37) \qquad \begin{aligned} \lim_{n,k \to \infty} V\left(\frac{n}{2k}, \frac{n\epsilon_n}{k}, k\right) &= \lim_{n,k \to \infty} V\left(\frac{n}{2k}, \frac{Z}{\sqrt{k}} \sqrt{\frac{n}{k}}, k\right) \\ &= 1 - \exp\left[-2\frac{n}{k} Z^2 - \sqrt{\frac{n}{k}}\left(\frac{n - 2k}{\sqrt{k}}\right) Z\right] \end{aligned}$$

so that an approximate evaluation of P_2 is given by

$$(38) \qquad P_2 = \sum_{k=[n(\frac{1}{2}+\Delta-\epsilon_n)]}^{n} B(k; n, \tfrac{1}{2}) \left[\exp\left(-\frac{2^2_n \epsilon^2_n}{k} + 2n\epsilon_n - \frac{n^2\epsilon^2_n}{k} \right) \right].$$

This formula was used to evaluate P_2 for a number of values of n and Δ. These are shown in Table 1. The striking feature of the table is the negligible size of P_2.

Further, by making the change of variable $W = 1 - U$ and noting that

$$(39) \qquad P_3 \leqq \Pr \left[\sup_{u_0+\Delta<u<1} [u - F_n(u)] \geqq \epsilon \mid u_0 - F_n(u_0) < \epsilon_n, G_{mu_0} \right]$$

it is obvious that for $u_0 = \frac{1}{2}$, $P_3 \leqq P_2$.

Hence $\min_{u_0} \beta_{D_n^-}(G_{mu_0})$ is bounded between $P_1 + P_2$ and $P_1 + 2P_2$ for large samples.

5. Tests related to D_n^- test. Anderson and Darling [1] considered a class of tests based on the more general distance function

$$\sup_{-\infty<x<\infty} \sqrt{n} \mid F_n(x) - F(x) \mid \psi[F(x)],$$

where ψ is a non-negative weight function. The choice of $\psi = 1$ yields the Kolmogorov statistic. Anderson and Darling also studied

$$\psi(t) = \begin{cases} \dfrac{1}{t(1-t)}, & 0 < a \leqq t \leqq b < 1, \\ 0, & \text{otherwise}, \end{cases}$$

but the distribution function is not in usable form. The distribution of $\sup_{-\infty<x<\infty}\sqrt{n}\,[F(x) - F_n(x)]\psi[F(x)]$, when H_0 is true, has apparently only been obtained for the case $\psi = 1$. More recently, Pyke [17] has studied a class of tests based on a generalized one-sided distance, but again the distributions have not been given.

TABLE 1

$P_2 = \Pr [\sup_{0\leqq u<u_0} u - F_n(u) \geqq \epsilon_n \mid u_0 + \Delta - F_n(u_0 + \Delta - 0) < \epsilon_n]$*

Δ	n			
	50	100	200	400
0.05	.0081	.0062	0035	.0011
0.10	.0027	.0001	—	—
0.20	.0002	—	—	—
0.30	—	—	—	—
0.40	—	—	—	—
0.50	—	—	—	—

* Calculations made using formula (38). Entries marked with — are less than .0001

One asymptotic result of this type is known that could form the basis of a la sample test of H_0. This is the result due to Renyi [18], viz., if H_0 is true

$$\lim_{n\to\infty} P\left\{\sqrt{n}\sup_{a\leqq u}\frac{[F_n(u) - u]}{u} < Z\right\} = \sqrt{\frac{2}{\pi}}\int_0^{Z[a/(1-a)]^{\frac{1}{2}}} e^{-t^2/2}\,dt, \quad Z >$$
$$= 0, \quad Z \leqq \tag{40}$$

for arbitrary a, $0 < a < 1$.

The restriction $a \leqq u$ is unpleasant since it imposes an additional decision the statistician, viz., the choice of a. Furthermore, it is apparent that the t based on this result cannot be consistent against alternatives which do not dif from $F_0(x)$ for the set $E[x:F_0(x) < a]$. On the other hand the test is consiste against all other alternatives in $\bar{\omega}$.

One feature of this test may be noted. The minimum power of the test may studied in a manner parallel to that used for the D_n^- test. In particular the pro ability of rejection is the probability that the empirical d.f. $F_n(u)$ falls at sor point below the line $u(1 + \epsilon_n) - \epsilon_n$ where ϵ_n is chosen to satisfy the size con tion; i.e., approximately for large samples

$$\epsilon_n = Z_\alpha\sqrt{\frac{1-a}{an}}. \tag{41}$$

The primary term of the power function $\beta(G_{mu_0})$ is thus seen to be appro imately

$$\Phi\left[\frac{\sqrt{n}\Delta - Z_\alpha\sqrt{\frac{1-a}{a}}(1 - u_0 - \Delta)}{[u_0(1 - u_0)]^{1/2}}\right], \tag{42}$$

which for sufficiently large n is minimized when

$$u_0 = 1 - a - \Delta.$$

Further this minimum power will be an increasing function of a; i.e., increasing will increase the minimum power of the test within the class of d.f.'s for which th test is consistent but at the same time this class will be decreased.

6. Tests based on the integral criterion. To Cramér and Von Mises is due th idea of testing H_0 by a statistic based on the integral of the square of the diffe ence between hypothetical and empirical distribution functions. Smirnov mod fied this by integrating with respect to the probability measure generated b $F(u)$. A more general form was given by Anderson and Darling [1] (this pape also gives references to the original authors which have been omitted here). This i

$$W_n^2 = n\int_{-\infty}^{\infty}[F_n(x) - F(x)]^2\psi[F(x)]\,dF.$$

The limiting distribution of this statistic with the weight function $\psi = 1$ wa given first by Smirnov, then by Von Mises and later by Anderson and Darling

They also gave a tabulation of the limiting distribution (cf. [1], p. 203). The latter authors also give the d.f. of W_n^2 for the weight function

$$\psi(t) = t(1 - t)$$

but the function is complex and no tabulation has been given

Before discussing the classical form of W_n^2, *it is of interest to note that since* we are here considering one-sided alternatives, it is not unreasonable to introduce as a test statistic

$$(43) \qquad W_n' = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]\, dF(x) = n \int_0^1 [F_n(u) - u]\, du.$$

It is seen at once that

$$(44) \qquad W_n' = \sum_{i=1}^{n} U_i - \frac{n}{2}$$

so that the test is equivalent to one based on

$$(45) \qquad \bar{U} = \frac{1}{n} \sum_{i=1}^{n} U_i .$$

Such a test has also been proposed by L. Moses.

For n large, under H_0, $\bar{U}$ is $N(\frac{1}{2}, 1/12n)$, while under th alteernative $G(u) < u$ it is normally distributed with mean $\int_0^1 u\, dG(u) > \frac{1}{2}$. The variance of $\bar{U}$ under the alternative is finite so that the test is consistent for all alternatives in $\bar{\omega}$. The test is also obviously monotone and hence p.o. For any alternative the large sample power is easily computed. In particular

$$(46) \qquad \bar{\beta}_{\bar{U}}(\Delta) = \beta_{\bar{U}}(G_M) = 1 - \Phi\left(\frac{Z_\alpha - \sqrt{3n}\Delta(2 - \Delta)}{[1 - \Delta^2(6 - 8\Delta + 3\Delta)]^{1/2}}\right)$$

and

$$(47) \qquad \beta_{\bar{U}}(G_{mu_0}) = 1 - \Phi\left(\frac{Z_\alpha - \sqrt{3n}\Delta^2}{[1 - \Delta^2(6 - 8\Delta + 3\Delta^2) + 12u_0\Delta^2]^{1/2}}\right).$$

The minimum of (47) is attained when $u_0 = 0$.

Consider now the classical integral criterion, i.e.,

$$(48) \qquad \omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2\, dF(x) = \int_0^1 [F_n(u) - u]^2\, du.$$

It is well known that

$$(49) \qquad \begin{aligned} n\omega^2 &= \frac{1}{12n} + \sum_{i=1}^{n} \left[U_i - \frac{2i - 1}{2n}\right]^2 \\ &= \sum_{i=1}^{n} U_i^2 - \frac{1}{n} \sum_{i=1}^{n} U_i(2i - 1) + \frac{n}{3} \end{aligned}$$

and that if H_0 is true

$$E(\omega^2) = \frac{1}{6n}, \qquad \sigma^2(\omega^2) = \frac{1}{n^2}\left(\frac{4n-3}{180n}\right). \tag{50}$$

It is known that if H_0 is true $n\omega^2$ has a limiting distribution which is not normal. However, if H_0 is false, the limiting distribution of ω^2, appropriately normalized, is normal.

For if the U_i have a d.f. $G(u)$

$$\omega^2 = \int_0^1 [u - G_n(u)]^2\, du = \int_0^1 [u - G(u) + G(u) - G_n(u)]^2\, du$$

$$= \int_0^1 \delta^2(u)\, du + 2\int_0^1 \delta(u)G(u) \tag{51}$$

$$- 2\int_0^1 \delta(u)G_n(u) + \int_0^1 [G(u) - G_n(u)]^2\, du,$$

where $G_n(u)$ has been written to emphasize that the sample has been drawn from the population with distribution G and where we have written $u - G(u) = \delta(u)$.

The notation

$$\int_0^u \delta(t)\, dt = D(u)$$

and

$$\int_0^1 \delta^2(u)\, du + 2\int_0^1 \delta(u)G(u) - 2D(1) + 2E[D(U)] = C(G)$$

will also be used.

From Kolmogorov's theorem that

$$\lim_{n\to\infty} [\Pr [\sqrt{n} \sup_{0<u<1} |G(u) - G_n(u)| \geqq Z]] = 2\sum_{v=1}^{\infty} (-1)^{v-1} e^{-2v^2 Z^2} \tag{52}$$

it is easily seen that

$$\sqrt{n}\int_0^1 [G(u) - G_n(u)]^2\, du$$

tends to zero in probability. Also

$$\int_0^1 \delta(u)G_n(u)\, du = \frac{1}{n}\sum_{i=1}^{n-1} i \int_{U_i}^{U_{i+1}} \delta(u)\, du$$

$$= D(1) - \frac{1}{n}\sum_{i=1}^{n} D(U_i). \tag{53}$$

Since $D(u) \leqq \frac{1}{2}$ for $0 \leqq u \leqq 1$, $E_G[D(U)]^2 < \infty$ and hence

$$\sqrt{n}\,\{1/n \textstyle\sum_{i=1}^n [D(U_i) - E[D(U)]]\}$$

is asymptotically normal with mean zero and variance given by the usual formula.

Finally then $\sqrt{n}\,(\omega^2 - C(G))$ is the sum of an asymptotically normal r.v. and one tending in probability to zero. It is therefore itself asymptotically normal with expectation zero.

Define ω_α by the equation $\Pr[n(\omega)^2 > \omega_\alpha \mid H_0] = \alpha$.

The ω^2 test (i.e., reject H_0 when $n\omega^2 > \omega_\alpha$) is consistent but not monotone. Its failure to be monotone arises from the fact that the test is two-sided and we are here considering one-sided alternatives. On the other hand, at least for n sufficiently large that the term

$$\int_{-\infty}^{\infty} [G(u) - G_n(u)]^2\, du$$

is negligible with respect to the other terms of ω^2, the test is p o This follows from the decomposition (51), since the other terms in this expression increase as G decreases.

The calculation of $E[D(U)]$, $\sigma^2[D(U)]$ is particularly simple for the alternatives G_{mu_0} and G_M. In fact, it is also possible in these cases to calculate straightforwardly $E(\omega^2)$.

Thus

$$E_{G_{mu_0}}(\omega^2) = \frac{\Delta^3}{3}\left(1 + \frac{1}{n}\right) + \frac{1}{n}\left(\tfrac{1}{6} + \Delta^2[u_0 - \tfrac{1}{2}]\right) \tag{54}$$

while

$$\sigma^2_{G_{mu_0}}(\omega^2) = \frac{u_0(1-u_0)}{n}\,\Delta^4 + 0\left(\frac{1}{n^2}\right) \tag{55}$$

The value of u_0 which minimizes the function $\beta(G_{mu_0})$ is a rather complicated expression involving Δ, n and ω_α; however, it is easily seen that as $n \to \infty$ this minimizing value tends to $\frac{1}{2}$. For simplicity we have evaluated only the approximate large sample power function $\beta(G_{m,\frac{1}{2}})$:

$$\beta_{\omega^2}(G_{m,\frac{1}{2}}) = 1 - \Phi\left[\frac{2}{\Delta^2}\left(\frac{\omega_\alpha^2}{\sqrt{n}}\right) - \frac{2\Delta}{3}\left(1 + \frac{1}{n}\right)\sqrt{n} - \frac{1}{3\Delta^2\sqrt{n}}\right]. \tag{56}$$

Similarly evaluating $E(\omega^2)$ and $\sigma^2(\omega^2)$ to terms of order $1/n^2$

$$\bar{\beta}_{\omega^2}(\Delta) = \beta_{\omega^2}(G_M) = 1 - \Phi(x), \tag{57}$$

where

$$x = \frac{\dfrac{\omega_\alpha - \left(\dfrac{1}{6} - \dfrac{\Delta^2}{2} + \dfrac{\Delta^3}{3}\right)}{\sqrt{n}} - (\Delta^2 - \tfrac{2}{3}\Delta^3)\sqrt{n}}{2\Delta\left[\dfrac{1}{12} - \dfrac{\Delta^2}{2} + \dfrac{2}{3}\Delta^3 - \dfrac{\Delta^4}{4}\right]^{1/2}} \tag{58}$$

7. Other tests. A procedure that has been suggested for the problem of combining tests and which consequently could be adapted to the equivalent problem of testing H_0, is based on the minimum or maximum of the transformed observa-

tions, i.e., in our notation, U_1 or U_n. Even restricting the problem by choosing a simple univariate statistic such as U_1, does not yield a unique u. m. p. test. Moreover the "intuitive" test of H_0 against one-sided alternatives—i.e., reject H_0 when $U_1 > c$ for appropriately chosen c—is obviously not consistent. In fact, it is only consistent for those alternatives $G(u) < u$ such that $\lim_{u \to \infty} G(u)/u = 0$. Furthermore, the test—reject H when $U_n > c$—would be consistent for no alternatives of $\tilde{\omega}$.

Of more interest are a group of tests based on another class of statistics, the so-called spacing of the observations. It is convenient to define

$$S_i = U_i - U_{i-1}, \qquad i = 1, 2, \cdots n + 1. \tag{59}$$

Various tests based on the statistics S_i have been proposed by Sherman [19] and others. These tests are not p.o. and hence are excluded from the present study. The proof of this fact as well as some other properties of these tests will be given in a later paper.

TABLE 2A

Minimum power of several tests for alternatives whose distance from F_0 is Δ

Test	Δ										
	0.05	0.1	0.125	0.15	0.175	0.20	0.25	0.3	0.35	0.45	0.50
						$n = 50$					
π, π'	.052	.059	—	—	—	.072	—	.108	—	.179	.306
$\bar{U}$	.052	.059	.065	.073	.085	.102	.153	.250	.577	.676	.994
ω^2						.131	.448	.697	.842	.922	.981
D_n^-	.057	.156	.248	.372	.511	.648	.862	.964	—	1.000	
						$n = 100$					
π, π'	.053	.063	—	—	—	.080	—	.137	—	.254	.460
$\bar{U}$	.053	.065	—	.092	—	.148	.257	.457	.738	.949	1.000
ω^2				.054	.228	.449	.730	.914	.970	.990	.999
D_n^-	.086	.327	.521	.710	—	.940	—	1.000			
						$n = 200$					
π, π'	.054	.068	—	—	—	.094	—	.185	—	.382	.623
$\bar{U}$	.055	.075	—	.123	—	.232	.447	.756	.965	1.000	
ω^2			.069	.335	.617	.803	.956	.991	—	1.000	
D_n^-	.158	.649	.862	.964	—	.999	1.000				
						$n = 400$					
π, π'	.056	.076	—	—	—	.117	—	.270	—	.583	.902
$\bar{U}$	.058	.091	.125	.179	.264	.388	.727	.966	—	1.000	
ω^2	—	.054	.415	.757	.916	.974	.998				
D_n^-	.329	.940	—	1.000							

8. Comparison of the minimum and maximum powers of consistent, partially ordered tests. In the preceding sections it has been shown that the tests associated with the statistics $\pi(8)$, $\pi'(9)$, D_n^- (24), $\bar{U}$ (45) and ω^2 (48) are consistent, monotone and p.o Furthermore, useful large sample approximations were found for $\underline{\beta}(\Delta)$ and $\bar{\beta}(\Delta)$ for each test. In view of the fact that most of these large sample power functions are expressed in terms of normal probabilities it would not be difficult to obtain inequalities between the power functions for the different tests. However, in not all cases does the same relationship between the power functions persist for all Δ or all n. Furthermore, such inequalities do not indicate the magnitude of the power differences.

As a more informative approach calculations have been made of $\underline{\beta}(\Delta)$ and $\bar{\beta}(\Delta)$ for each test for a range of values of n and Δ. These have been calculated for two test sizes, viz., $\alpha = 0.05$ and $\alpha = 0.01$. The minimum power was calculated for $\Delta = 0.05, 0.1, 0.2, 0.3, 0.4$ and 0.5 and where desirable, some intermediate values while the maximum power was calculated for $\Delta = 0.01$, (0.01) 0.10, 0.15, 0.20, 0.30, 0.40, and 0.50. A fixed sequence of sample sizes n was used, viz., $n = 50, 100, 200, 400, 600, 800, 1000, 2000, 4000, 6000, 8000, 10{,}000 \ldots$ with the stopping rule, stop whenever the absolute values of the normal deviate exceeded 3.

TABLE 2B

Maximum power of several tests for alternatives whose distance from F_0 is Δ

Δ	0 01	0 02	0 03	0 04	0 05	0.06	0 07	0 08	0 09	0 10	0 125	0 15	0 2
						$n = 50$							
π'	959	.965	.972	.977	.983	.987	.990	.993	995	997	1		
π	.080	.125	.189	.272	.373	484	.598	706	.799	.871	970	.996	
$\bar{U}$	.080	.123	.178	.246	.325	412	.504	.595	.681	758	.899	.968	.999
ω^2	—	—	—	.092	.193	.308	412	509	.589	.662	—	—	.973
D_n^-	.070	.096	129	.170	.220	.278	.346	.420	501	.586	.793	948	
						$n = 100$							
π	.967	.980	.989	.949	.998	1							
π'	.111	.211	.354	.523	.689	.824	.916	.966	.988	.997	1		
$\bar{U}$	.096	.168	267	.387	.518	.646	759	.849	914	955	.994	1	
ω^2	—	—	.118	.277	.423	.562	.668	.754	.818	.869	—	—	999
D_n^-	.080	.123	.181	.257	.350	.459	.578	.698	.811	.905	1		
						$n = 200$							
Δ	0.01	0 02	0 03	0 04	0 05	0 06	0 07	0 08	0.09	1 0			
π	.169	.370	.633	.840	.950	.989	.999	1					
$\bar{U}$	.124	.250	.421	.609	.773	.889	.954	.984	.996	.999			
ω^2	—	.093	.317	.529	.675	.797	.873	.925	.955	.975			
D_n^-	.096	.171	.279	.421	.587	.755	.897	.983	1				

the powers so calculated for $\alpha = 0.05$ are exhibited in Table 2A and it might be hoped that some of the tests could be eliminated by such a comparison—this would be the case if $\bar{\beta}$ for one test fell below $\underline{\beta}$ for some other test. However, this is not the case.

In general the tables indicate that the relationship of the tests is reversed from the minimum to the maximum power. Thus we have

$$\underline{\beta}_{\pi} = \underline{\beta}_{\pi'} < \underline{\beta}(\omega^2), \qquad \underline{\beta}_{\bar{U}} < \underline{\beta}_{D_n^-} < \bar{\beta}_{D_n^-} < \bar{\beta}(\omega^2), \qquad \bar{\beta}_{\bar{U}} < \bar{\beta}_{\pi} < \bar{\beta}_{\pi'} .$$

The relationship between the ω^2 and $\bar{U}$ tests varies with Δ and n.

It is evident that the π' test has the best maximum power of the tests considered, but its minimum power (and that of the π test) is extremely low. On the other hand the D_n^- test which has the lowest maximum power (of the tests considered) has the greatest minimum power. This raises the question whether there exists a non-trivial test which is p.o. and for which $\underline{\beta}(\Delta) = \bar{\beta}(\Delta)$.

An alternative comparison between the tests is given in Table 3, which shows the sample sizes necessary to achieve a pre-assigned power level β, for given Δ and for $\alpha = 0.05$. The values corresponding to $\beta = 0.95$ only are listed though corresponding values of n have been calculated for $\beta = 0.90$ and $\beta = 0.99$. The latter calculation emphasizes the poorness of the π, π' tests against alternatives G_{mu_0}—over 2,443,900 observations are required to insure $\underline{\beta}(0.05) = 0.99$. It should be noted that these values of n were calculated from the primary term

TABLE 3

Sample sizes necessary that $\underline{\beta}(\Delta)$ *and* $\bar{\beta}(\Delta) = 0.95$ *for several* p.o. *tests and for* $\alpha = 0.05$

Minimum Alternative

Test	Δ					
	0.05	0.1	0.2	0.3	0.4	0.5
D_n^-	1675	419	105	47	27	17
ω^2	14,038	2290	406	153	78	45
$\bar{U}$	569,067	34,233	1867	304	77	25
π, π'	1,677,025	102,081	23,903	4463	1325	511

Maximum Alternative

Test	Δ									
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
D_n^-	29,679	7420	3298	6855	1188	825	606	464	367	297
ω^2	4761	1057	540	302	204	160	104	80	65	53
$\bar{U}$	9108	2296	1027	583	375	261	193	148	117	95
π	3067	936	471	291	200	148	115	92	77	65

P_1[cf(29)] and consequently the required sample size with the D_n^- test is slightly over-estimated.

It is also to be noted that the smaller sample sizes indicated in Table 3 must not be construed too literally since they have been computed from asymptotic formulae.

Of these tests considered it appears that if no information is available on the possible alternatives to H_0 then from some minimax point of view, the D_n^- test is the most favorable.

Acknowledgment. The calculations on which Tables 1, 2, and 3 are based were made by Mrs. P. Freeman; some were also made by Mr. P. Stevens. Funds for these calculations were provided by the Agnes Anderson Fund of the University of Washington, for which the author is extremely grateful.

REFERENCES

[1] T W. ANDERSON AND D. A. DARLING, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math Stat.*, Vol. 23 (1952), pp. 193-212.

[2] ALLAN BIRNBAUM, "Combining independent tests of significance," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 559-574

[3] Z. W. BIRNBAUM AND F. TINGEY, "One-sided confidence contours for probability distribution functions," *Ann Math Stat*, Vol. 22 (1951), pp. 592-596.

[4] Z. W. BIRNBAUM, "Distribution free tests of fit for continuous distribution functions," *Ann. Math. Stat*, Vol. 24 (1953), pp 1-8

[5] Z W. BIRNBAUM, "On the power of a one-sided test of fit for continuous distribution functions," *Ann. Math. Stat*, Vol. 24 (1953), pp. 284-489.

[6] Z. W. BIRNBAUM AND H. RUBIN, "On distribution free statistics," *Ann Math. Stat.*, Vol. 25 (1954), pp. 593-598.

[7] Z. W. BIRNBAUM AND ERNEST M. SCHEUER, *On the Power of a One-sided Test of Fit Against Stochastically Comparable Alternatives*, Technical Report No 15, O N R. Contract, Laboratory of Statistical Research, University of Washington.

[8] Z. W. BIRNBAUM AND RONALD PYKE, *On Some Distributions Related to the Statistic* D_n^+, Technical Report No 23, O N.R Contract, Laboratory of Statistical Research, University of Washington

[9] G. E. P. BOX AND S. L. ANDERSEN, "Permutation theory in the derivation of robust criteria and the study of departures from assumption," *J. Roy Stat. Soc.* (Series B), Vol. 17 (1955), pp. 1-26.

[10] D. A. DARLING, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239-253

[11] MONROE D. DONSKER, "Justification and extension of Dorb's heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat*, Vol. 23 (1952), pp. 277-281.

[12] J. L. DOOB, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann Math. Stat.*, Vol. 20 (1949), pp. 393-403.

[13] W. FELLER, *An Introduction to Probability Theory and Its Application*, John Wiley & Sons, New York, 1950.

[14] J. L. HODGES AND E. L. LEHMANN, "The efficiency of some nonparametric competitors of the t-test," *Ann. Math Stat.*, Vol. 27 (1956), pp. 324-335.

[15] E L. LEHMANN, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24 (1956), pp. 23-43.

[16] P. Moran, "The random division of an interval," *J. Roy. Stat. Soc. Suppl.*, Vol. 9 (1947), pp. 92–98.
[17] R. Pyke, "On one-sided distribution free statistics," Unpublished thesis, Dept. of Mathematics, University of Washington, 1956.
[18] A. Renyi, "On the theory of order statistics" *Acta. Math. Acad. Sci. Hung.*, Vol. 4 (1953), pp. 191–231.
[19] B. Sherman, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339–361.
[20] N. V. Smirnov, "On the deviation of the empirical distribution functions," *Rec. Math.* (*Mat. Sbornik*) (*NS*), Vol. 6 (1939), pp. 3–26.

ON THE NONRANDOMIZED OPTIMALITY AND RANDOMIZED NONOPTIMALITY OF SYMMETRICAL DESIGNS[1]

By J. Kiefer[2]

Cornell University

0. Summary. Many commonly employed symmetrical designs such as Balanced Incomplete Block Designs (BIBD's), Latin Squares (LS's), Youden Squares (YS's), etc., are shown to have optimum properties among the class of *non-randomized*[1] designs (Section 3). This represents an extension of a property first proved by Wald for LS's in [1]; a similar property demonstrated by Ehrenfeld for LS's in [2] (as well as a third optimum property considered here) is shown to be an immediate consequence of the Wald property, and the Wald property is shown to be the more relevant when one considers optimality rigorously (Section 2). Surprisingly, all of these optimum properties fail to hold if *randomized*[1] designs are considered (Section 4); the results of Sections 2 and 3, as well as those appearing previously in the literature (as in [1], [2], [3]) must be interpreted in this sense. Generalizations of the BIBD's and YS's, for which analogous results hold, are introduced.

1. Introduction. Wald [1] stated an optimality criterion (called E-optimality in Section 2) for designs used in testing hypotheses in the setting of two-way soil heterogeneity where LS's are commonly employed, and succeeded in proving that a slightly different criterion (called D-optimality in Section 2) is satisfied by the LS design. Wald also stated that an analogous result holds for Graeco-Latin Squares and higher Latin Squares This statement gives rise to speculation when one considers that, in a 3×3 Graeco-Latin Square (or, more generally, in an $n \times n$ square of order $n - 1$), there are no degrees of freedom for error: this implies that any test (e.g., of the hypothesis H_0 that there are no treatment effects) whose size (= supremum of the power function under H_0) is α, has a power function whose infimum over any of the contours usually considered ($\psi(\mu)/\sigma^2$ = constant, as discussed in the sequel) is $\leqq \alpha$. It is easy to construct a better design, i.e., one for which the infimum of the power function of some test over such a contour is > the size of the test; for example, for each of the two

Received July 8, 1957; revised January 22, 1958.

[1] One of the referees of this paper felt that the following remark on nomenclature should be included: Throughout this paper, the term *randomized design* is used in describing a statistical procedure which chooses according to a prescribed probability mechanism a member of a given class of ordinary designs, the chosen design being the one actually used; a precise definition is given in the text. The properties of such a procedure take into account the probabilities of the various possible choices. A *nonrandomized* design chooses one member of the given class with probability one. The customary usage of the phrase *randomized design* in the design of experiments can be viewed as a special case of the decision-theoretic usage employed here, but the reader is warned not to interpret the phrase in that narrower sense.

[2] Research sponsored by the Office of Naval Research.

factors, with probability $\frac{1}{2}$ use an ordinary LS design on the three levels of that factor holding the level of the other factor fixed.[3]

The phenomenon just described makes one wonder whether the optimality result for ordinary LS's also fails to hold if one permits comparison with randomized designs.[1] At the same time, the question arises whether an analogue of the limited optimality property of the LS (or Graeco-LS) design holds in a wide class of design settings for designs with suitable symmetry properties, and whether these designs fail to be optimum when compared with randomized designs.[1] This paper answers these questions affirmatively.

In Section 2A we define four optimality criteria (D-, E-, M-, and L-optimality) for designs (especially, for the normal case); Wald [1] and Ehrenfeld [2] proved D- and E-optimality, respectively, for the LS design. It is indicated why M-optimality, the strongest and least artificial of the four, seems very difficult to verify in most problems (although L-optimality, which is a local version of M-optimality, can sometimes be verified). At the same time, we list briefly for later reference the known results on the Analysis of Variance Test which are used in optimality considerations, and point out the incorrectness of tacitly assuming (as previous work in this area has done) that one should use that test, whatever design is chosen. In Section 2B we indicate by example why E-optimality seems, at least in the present state of knowledge indicated in 2A, the least satisfactory of the criteria considered; the connection of D-optimality with Isaacson's notion of type D tests [11] is examined. In Section 2C it is shown in a general setting where there is suitable symmetry that D-optimality implies E-optimality and L-optimality.

In Section 3A it is indicated why the treatment of LS's is much simpler than that of YS's, BIBD's, etc., and the general treatment of incomplete block designs

---

[3] It should be evident that the example of the 3 × 3 Graeco-Latin square, as well as the example discussed in the fourth paragraph below wherein two observations are taken, are of no *practical* importance; these simple examples are given to illustrate the general principles of Section 4. Those principles show that a precise study of certain optimality criteria for designs associated with familiar problems of testing hypotheses, can lead to the unexpected conclusion that certain intuitively unappealing randomized designs are superior to certain intuitively appealing nonrandomized symmetrical designs. The principles are less transparent (although applicable) in the context of applicationally meaningful problems such as those of Section 4, than in the simple examples; hence, the latter examples are discussed first. The present comments are included because two referees apparently read these simple examples as practical suggestions. In the same light, it is clear that the design δ in the fourth paragraph below, as well as its analogues in Section 4, is not suggested to the practical worker who wants *estimates* of all treatment effects; for these designs illustrate a non-optimality property of classical nonrandomized symmetrical designs in *hypothesis testing*, and a local property at that (see Section 5.4). In fact, the results of Section 4 are not even relevant for most estimation problems (see Section 5.2). To the practical worker who objects (as at least one has) to the conclusions of Section 4 on the grounds that one should not use a design which does not estimate all treatment effects, it should be pointed out that (1) the classical nonrandomized symmetrical design may still possibly possess certain *global* optimality properties (see Section 5.4), and (2) perhaps his problem is not really one of testing hypotheses.

of Bose [4] is briefly recalled; this treatment proves more useful in Section 3C than the more direct least squares approach used in [1] and [2] would be. In Section 3B several algebraic propositions (emphasizing the role of symmetry) are verified, which can be used to prove D- and E-optimality in important examples. Several such examples are considered in Section 3C, including generalizations of the BIBD's and the YS's.

Section 4 contains two theorems the consequences of which are that nonrandomized symmetrical designs are not optimum if randomization is permitted. In Section 4B it is shown that, whether or not the variance is known, for α sufficiently small there is a randomized design whose power function is uniformly larger than that of the symmetrical design in some neighborhood of the hypotheses H_0 that all treatment effects are the same. This is slightly less transparent than the result of Section 4A, which gives an analogous result for *all* α when the above H_0 is replaced by the hypothesis that all treatment effects are equal to some specified value. The latter result can best be understood by considering the simplest example[3]: Suppose X_{ij} normal with unit variance and mean μ_i and that all X_{ij} are independent ($i, j = 1, 2$). Our problem is to select (before observation) exactly two of the X_{ij} and use them to test $\mu_1 = \mu_2 = 0$ against some class of alternatives. The symmetrical design d(say) selects X_{11} and X_{21} and uses the usual χ^2 test, and obviously has constant power $> \alpha$ on the contour $\mu_1^2 + \mu_2^2 = c > 0$, while either of the designs d_i ($i = 1, 2$), where d_i uses X_{i1} and X_{i2}, has α for the infimum of the power function on this contour. Let δ be the randomized design[1] obtained by using d_1 or d_2 with probability $\frac{1}{2}$ each. It is easily seen that, for μ_1 and μ_2 near 0, the power function of δ is $\alpha + c_1(\mu_1^2 + \mu_2^2) +$ terms of higher order, where $c_1 > 0$. Thus, on the contour $\mu_1^2 + \mu_2^2 = c > 0$ with c small, the power function of δ is almost constant and hence approximately equal to the value at $\mu_1 = \mu_2 = (c/2)^{\frac{1}{2}}$. Thus, in comparing d and δ near H_0, we may to a first approximation assume $\mu_1 = \mu_2$. But δ is clearly optimum for testing $\mu_1 = \mu_2 = 0$ *assuming* $\mu_1 = \mu_2$, while d (whose test is based on $X_1^2 + X_2^2$) is not. This explains why, for c small, δ has a power function greater than that of d.

Many of the results of this paper have counterparts for problems of point and interval estimation, for other distributions, etc. Such extensions and generalizations, as well as various other remarks, are stated in Section 5.

In design settings where no suitably symmetric design exists, it is often tedious algebraically to show that a design which is "closest to symmetrical" is optimum (if it *is* optimum: see the example of Section 2B), and we omit such considerations here. On the other hand, the conclusions of Section 4 have little to do with whether or not symmetrical designs are being considered.

Throughout this paper, except where explicitly stated to the contrary, Y will denote an N element column vector whose components Y_i are independent normal random variables with common variance σ^2 (it will be explicitly stated whenever σ^2 is assumed known; whether or not σ^2 is known has very little effect on our results); μ is an unknown m-vector, X_d is a known $N \times m$ matrix depending on an index d (the "design") and which will be described further below.

and the expected value of Y when μ and σ^2 are the parameter values and whei the design d is used is

$$E_{\mu,\sigma;d}Y = X_d\mu. \tag{1.1}$$

X_d is, within limits, subject to choice by the experimenter. (In many application it is a matrix of zeros and ones.) We denote by Δ the set of choices of the index which are available to the experimenter. A randomized design[1] δ is a probabilit measure on Δ (the latter will usually be finite in this paper, and measurabilit considerations will be trivial otherwise) which is used by selecting a d from according to this measure and then using the selected d. We denote the class o available δ by Δ_R.

In many problems, one imposes an additional assumption of the form $\Gamma\mu = \gamma$ where Γ and γ are known $g \times m$ and $g \times 1$ matrices. Such an assumption can be absorbed into (1.1) and we suppose this to have been done, with no loss o generality.

A hypothesis H_0 will in this paper be of the form $R\mu = 0$, where R is a specifie $r \times m$ matrix ($r \leqq m$) which we can take to be of rank r with no loss of gener- ality. For simplicity, we can think of the class H_1 of alternatives as being all for which $R\mu \neq 0$. (For simplicity, we assume that σ^2 is either known exactly o else is known only to be positive, under both H_0 and H_1.) A hypothesis of the form $R\mu = \rho$ is easily reduced to the above form by letting p satisfy $Rp =$ and replacing Y by $Y^* = Y - X_d p$ and μ by $\mu^* = \mu - p$ in (1.1).

We introduce some notation to be used in Section 2. We denote the $k \times k$ identity matrix by I_k. The transpose of a matrix A is written A'. It may or may not be that all r elements of $R\mu$ are estimable when a given design d is used. Suppose that there are s_d linearly independent linear combinations of the elements of $R\mu$ which have unbiased estimators when d is used, but not $s_d + 1$ such combinations. Then there is an $s_d \times r$ matrix Q_d such that there exist linear unbiased estimators of all components of $Q_d R\mu$ when design d is used; let t_d be the s_d-vector of such estimators with minimum variance ("best linear estimators" or b.l.e.'s), and let $\sigma^2 V_d$ be the convariance matrix of the components of t_d. When $s_d = r$, we may take Q_d to be the identity; for this choice of Q_d, we shall denote V_d by $\bar{V}_d$. Let b_d be the rank of X_d. Then there are b_d linearly independent combinations of the components of μ which are estimable when d is used. Of these, s_d of them can be taken to be the elements of $Q_d R\mu$; thus, there exists a $(b_d - s_d) \times m$ matrix J_d of rank $b_d - s_d$ whose rows are orthogonal to those of $Q_d R$ (i.e., $J_d' Q_d R = 0$) and such that all components of $J_d\mu$ have unbiased estimates when d is used. Let L_d be the $b_d \times m$ matrix whose first $b_d - s_d$ rows are J_d and whose last s_d rows are $Q_d R$. Let $\bar{S}_d$ be the usual best unbiased estimator of σ^2 (if it is unknown), so that $(N - b_d)\bar{S}_d/\sigma^2$ has the χ^2-distribution with $h_d = N - b_d$ degrees of freedom (it may be that $h_d = 0$ and there is no $\bar{S}_d$). For any test ϕ_d associated with d, let $\beta_{\phi_d}(\mu, \sigma^2)$ be the power function of ϕ_d (of course, β_{ϕ_d} actually depends on μ only through $L_d\mu$). For $0 < \alpha < 1$ we denote by

$H_d(\alpha)$ the class of all ϕ_d of size α, i.e., all ϕ_d for which

$$\beta_{\phi_d}(\mu, \sigma^2) \leqq \alpha \text{ whenever } R\mu = 0; \tag{1.2}$$

and by $H_d^*(\alpha)$, the class of similar tests of size α, i.e., those for which (1.2) holds with the inequality sign replaced by equality. Finally, let $F_{d,\alpha}$ denote the usual F-test of H_0 of size α with s_d and h_d degrees of freedom, based on $t_d'V_d^{-1}t_d/s_d\bar{S}_d$ (if σ^2 is known, this is replaced by the appropriate χ^2-test).

The symbol $g_{i,j}(\alpha)$ is used to denote the derivative at H_0 of the power function of the F-test of size α and i, j degrees of freedom, with respect to (a common choice of) the parameter on which it depends; specifically, if $r = m = i$, $N - r = j$, the matrices R, Q_d, and V_d are the identity, and the true values of μ and σ^2 are such that $\mu'\mu/\sigma^2 = \lambda$, then, as $\lambda \to 0$, the power function of $F_{d,\alpha}$ is

$$\alpha + g_{i,j}(\alpha)\lambda + O(\lambda^2). \tag{1.3}$$

The results of this paper can be stated in a very general setting involving invariance of Δ, of the restriction $R\mu = 0$, and of a generalization of the function ψ considered below, as well as of certain designs, under an appropriate group of permutations of the components of μ. However, in order to make our proofs (and, in particular, the role of symmetry) as transparent as possible, we will carry them out in two cases; the reader will not find it difficult to state our results more generally by making appropriate linear transformations, etc. The two cases (Δ and X_d being further specified in particular examples, the role of the function ψ which distinguishes contours on which the power function is examined, will be seen in Section 2A) are:

Case I: $$\psi(\mu) = \sum_1^u \mu_i^2 \text{ and } R = R_I,$$

Case II: $$\psi(\mu) = \sum_1^u (\mu_i - \bar{\mu})^2 \text{ and } R = R_{II};$$

here we have written $\mu' = (\mu_1, \cdots, \mu_m)$, and $\bar{\mu} = \sum_1^u \mu_i/u$, while R_I is the $u \times u$ identity followed by $m - u$ columns of zeros (so $R_I\mu = 0$ means $\mu_1 = \cdots = \mu_u = 0$), and R_{II} is a $(u - 1) \times u$ matrix P followed by $m - u$ columns of zeros, where P consists of the last $u - 1$ rows of a $u \times u$ orthogonal matrix $\bar{O}$ whose first row elements are all $1/\sqrt{u}$ (so $R_{II}\mu = 0$ means $\mu_1 = \cdots = \mu_u$). The optimality results which hold in Case I are usually much more trivial to obtain than those of Case II, and Section 3B will therefore be mainly devoted to results applicable to the latter case, it being clear how to obtain the corresponding results in the former case.

2. Optimality criteria.

2A. *Preliminaries.* For a fixed design d, the test $F_{d,\alpha}$ is known to have several optimum properties, which we now list (there are obvious analogues when σ^2 is known):

(a) If $s_d = 1$ (and only then), among tests in $H_d(\alpha)$ which are unbiased (thi implies that the tests are in $H_d^*(\alpha)$), $F_{d,\alpha}$ is uniformly most powerful (UMP' See [5] (a trivial completeness argument characterizing similar tests is all that i required to allow the $J_d\mu$ which is not present in [5] to be introduced, carryin through the argument there for each fixed value of the b.l.e. of $J_d\mu$).

(b) Among tests in $H_d(\alpha)$, $F_{d,\alpha}$ is UMP invariant (under the usual group c transformations when the problem is reduced to canonical form). See [5].

(c) (Wald's theorem) Among tests in $H_d^*(\alpha)$, for each $c > 0$, $\sigma^2 > 0$, an value of $J_d\mu$, the test $F_{d,\alpha}$ maximizes the Lebesgue integral of $\gamma_{\phi_d}(\nu, J_d\mu, \sigma^2)$ o the sphere $\nu'\nu = c$, where $\nu = G_dQ_dR\mu$ with G_d nonsingular $s_d \times s_d$ is such tha the b.l.e.'s of the components of ν have σ^2 times the identity for their covarianc matrix (i.e., ν is the vector of parameters about which H_0 is concerned in th canonical form of the problem), and where $\gamma_{\phi_d}(G_dQ_dR\mu, J_d\mu, \sigma^2) = \beta_{\phi_d}(\mu, \sigma^2$ See [6] or [7] (the parenthetical remark at the end of (a) is relevant to [7] here

(d) (Hsu's theorem, a consequence of (c)) Among tests in $H_d(\alpha)$ whose powe function depends only on $\lambda_d = \mu'R'Q_d'V_d^{-1}Q_dR\mu/\sigma^2$ (this implies that the tests ar in $H_d^*(\alpha)$), $F_{d,\alpha}$ is UMP. See [8].

(e) Among tests in $H_d(\alpha)$, $F_{d,\alpha}$ is minimax (over H_1) for a variety of weigh functions, e.g., any nonnegative function of the λ_d of (d); in particular, $F_{d,}$ maximizes the minimum power on the contour $\lambda_d = c$ for each $c > 0$. See [9] o [10] (the result follows from (c) if we restrict consideration to $H_d^*(\alpha)$).

(f) (A special case of (e)) $F_{d,\alpha}$ is most stringent in $H_d(\alpha)$. See [9] or [10].

(g) (A consequence of (c)) $F_{d,\alpha}$ is of type D in $H_d(\alpha)$. (See [11] or Section 2I below for definition of type D, and Section 2B for a proof.)

It is to be noted that all the above criteria of optimality of the test $F_{d,\alpha}$ ar relative to the design d. Thus, it is an error to assume (as has been done in pre vious papers on optimum designs) in a logical approach to optimum desig problems that one should automatically use the test $F_{d,\alpha}$, whatever the chosen d *when a reasonable criterion for optimality of a design, or of a test for a given design may dictate the use of a test other than* $F_{d,\alpha}$. In fact, the example of Section 2I really illustrates that the use of $F_{d,\alpha}$ need not lead to an optimum design or tes for many reasonable definitions of optimality; and the fact that it seems diffi cult (for many reasonable optimality criteria such as M-optimality, and fo many common design problems) to characterize the appropriate test, is wha makes it much harder than it has been thought to give a rigorous demonstratio of the optimality of various common designs. We now list four optimality cri teria for designs (there are many other obvious similar ones); the discussion o their meaning immediately follows the fourth definition.

M-optimality: For $c > 0$ and $0 < \alpha < 1$, a design d^* is said to be *$M_{\alpha,c}$-optimun in* Δ if, for some $\phi_{d^*}^*$ in $H_{d^*}(\alpha)$,

$$\inf_{\Gamma_c} \beta_{\phi^*_{d^*}}(\mu, \sigma^2) = \max_{d\epsilon\Delta} \sup_{\phi\epsilon H_d(\alpha)} \inf_{\Gamma_c} \beta_\phi(\mu, \sigma^2), \tag{2.1}$$

where Γ_c is the set of all μ, σ^2 for which $\psi(\mu)/\sigma^2 = c$.

L-optimality: A design is said to be *L_α-optimum in* Δ if, for some $\phi^*_{d^*}$ in $H_{d^*}(\alpha)$,

$$\lim_{c\to 0} [a_{\phi^*_{d^*}}(c) - \alpha]/[b(c) - \alpha] = 1, \tag{2.2}$$

where $a_{\phi^*_{d^*}}(c)$ and $b(c)$ are the expressions on the left and right sides of (2.1), respectively. A design is said to be *L-optimum in* Δ if it is L_α-optimum in Δ for $0 < \alpha < 1$.

D-optimality: A design d^* is said to be *D-optimum in* Δ if

$$\det \bar{V}_{d^*} = \min_{d\epsilon\Delta'} \det \bar{V}_d, \tag{2.3}$$

where Δ' is the set of d in Δ for which $s_d = r$, and if $d^* \epsilon \Delta'$.

E-optimality: A design d^* is said to be *E-optimum in* Δ if

$$\pi(\bar{V}_{d^*}) = \min_{d\epsilon\Delta'} \pi(\bar{V}_d) \tag{2.4}$$

and if d^* is a member of Δ', where $\pi(\bar{V}_d)$ is the maximum eigenvalue of $\bar{V}_d$.

The above definitions will also be used with Δ replaced by Δ_R. In that case, for any δ, $\bar{V}_\delta^{-1}$ is defined to be the expected value under δ of $\bar{V}_d^{-1}$, the latter being replaced by the inverse of the covariance matrix of the b.l.e. of the estimable components of $R\mu$ (with zeros adjoined to this inverse in appropriate places to make it $r \times r$) if $s_d < r$; Δ_R' is then the set of δ for which $\bar{V}_\delta^{-1}$ is nonsingular. (This $\bar{V}_\delta^{-1}$ appears in computing certain β_{ϕ_δ} near H_0.)

D-optimality and E-optimality have been discussed in [1] and [2] and will also be discussed in Section 2B, where it will be seen that they have to do with local properties (near H_0) or optimum properties *assuming the use of* $F_{d,\alpha}$. Unfortunately, $M_{\alpha,c}$-optimality in Δ (or, better, $M_{\alpha,c}$-optimality in Δ simultaneously for all c) seems very difficult to verify, even in many simple problems, although it does not require much temerity to conjecture that it holds in such cases as those discussed in Section 2C. A similar remark applies to L-optimality (see, however, Lemma 2.2), a local version (near H_0) of M-optimality. The source of this difficulty in verifying M-optimality is illustrated by the example of Section 2B; it is simply that for fixed d the test which achieves the supremum over ϕ on the right side of (2.1) need not be $F_{d,\alpha}$ and is generally hard to compute (as is therefore the right side of (2.1)).

2B. *D- and E- optimality.* We begin by describing the meaning of E-optimality (which criterion is stated in [1] and is verified for the LS design in [2]). Suppose for fixed α, that we agreed to restrict ourselves to using $F_{d,\alpha}$, whatever d is chosen. The power function of $F_{d,\alpha}$ is then a strictly increasing function of λ_d (defined in Section 2A(d)). Now, in either Case I or II, for any $c > 0$, if we want a design d for which $F_{d,\alpha}$ maximizes the minimum power on the contour $\psi(\mu)/\sigma^2 = c$ (i.e., *which is $M_{\alpha,c}$-optimum in Δ under the additional restriction that we use $F_{d,\alpha}$*), we may restrict our attention to Δ' (since, for $s_d < r$, the infimum of $\beta_{F_{d,\alpha}}$ on the contour $\psi(\mu)/\sigma^2 = c$ is α; if Δ' is empty, there is no problem). $F_{d,\alpha}$ has the same number of numerator degrees of freedom for all d in Δ'; if also

b_d is the same for each d in Δ' (this is often the case in important examples such as those of Section 3C) so that the denominator degrees of freedom are the same for all $F_{d,\alpha}$, then a design which maximizes the minimum power on $\psi(\mu)/\sigma^2 = c$ *simultaneously for all* c is precisely one which maximizes the minimum of λ_d subject to $\psi(\mu)/\sigma^2 = c$. Since $\psi(\mu) = (R\mu)'(R\mu)$ in both Cases I and II, this means maximizing $\min_{\xi'\xi=1} \xi'\bar{V}_d^{-1}\xi = 1/\pi(\bar{V}_d)$. This is precisely the criterion of E-optimality.

One can cite many practical examples to illustrate that the restriction to using $F_{d,\alpha}$, which is imposed in order to make E-optimality meaningful, can have serious detrimental consequences. The simplest possible situation will suffice as an example: Suppose $N > 2$, $r = m = 2$, $R = R_{\text{I}}$, and Δ' to consist of two designs with

$$\bar{V}_{d_1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \bar{V}_{d_2} = \begin{pmatrix} 1+\epsilon & 0 \\ 0 & \epsilon \end{pmatrix},$$

where $\epsilon > 0$. Clearly, d_1 is E-optimum. Moreover, if d_1 is used, optimum property (e) above states that, for every c, $F_{d_1,\alpha}$ maximizes the minimum power on the contour $(\mu_1^2 + \mu_2^2)/\sigma^2 = c$ among all tests in $H_{d_1}(\alpha)$. However, if d_2 is used, $F_{d_2,\alpha}$ does not have this property. For example, if d_2 is used, let ϕ' be the test which with probability $(1 + \epsilon)/(1 + 2\epsilon)$ uses the F-test (with 1 and N-2 degrees of freedom) of size α of the hypothesis $\mu_1 = 0$, and which with probability $\epsilon/(1 + 2\epsilon)$ uses the F-test of size α of the hypothesis $\mu_2 = 0$. The power function of ϕ' near $(\mu_1^2 + \mu_2^2)/\sigma^2 = 0$ is then

$$\alpha + g_{1,N-2}(\alpha)\ (\mu_1^2 + \mu_2^2)/(1 + 2\epsilon)\sigma^2 + o([\mu_1^2 + \mu_2^2]/\sigma^2),$$

while that of $F_{d_2,\alpha}$ is

$$\alpha + g_{2,N-2}(\alpha) \left(\frac{\mu_1^2}{1+\epsilon} + \frac{\mu_2^2}{\epsilon}\right) /\sigma^2 + o([\mu_1^2 + \mu_2^2]/\sigma^2).$$

The infimum of the expression multiplying $g_{2,N-2}(\alpha)$, taken on the contour $(\mu_1^2 + \mu_2^2)/\sigma^2 = c$, is $c/(1 + \epsilon)$, compared with $c/(1 + 2\epsilon)$ for the coefficient of $g_{1,N-2}(\alpha)$; since $g_{1,N-2}(\alpha)/g_{2,N-2}(\alpha) \to 2$ as $\alpha \to 0$ (see Lemma 4.3 below) the assertion three sentences above regarding $F_{d_2,\alpha}$ is verified. Moreover, since the power function of $F_{d_1,\alpha}$

$$\alpha + g_{2,N-2}(\alpha)(\mu_1^2 + \mu_2^2)/\sigma^2 + o([\mu_1^2 + \mu_2^2]/\sigma^2),$$

we see similarly that, at least for α, ϵ, and c sufficiently small, d_1 is *not* $M_{\alpha,c}$-optimum or L_α-optimum, ϕ' being locally uniformly more powerful than $F_{d_1,\alpha}$; thus, the assertion of the first sentence of this paragraph regarding E-optimality is verified.

Of course, for any fixed α, ϵ, and c we have not asserted that the test ϕ' (considered above only for illustrative purposes) is $M_{\alpha,c}$-optimum. If one uses d_2, the power functions of ϕ', $F_{d_2,\alpha}$, etc., are not constant on $(\mu_1^2 + \mu_2^2)/\sigma^2 = c$ (the same is true of the test which minimizes the integral of the power function on that contour), and the computation of the supremum over ϕ on the right side

of (2.1) does not seem easy (this will be discussed further in Section 5). Thus, the above example also illustrates why M-optimality (or L-optimality) seems so difficult to verify in many problems.

In order to see the meaning of D-optimality, we turn to the notion of a type D test as defined in [11] (we discuss the case where σ^2 is unknown, the other case being similar): For fixed d, let the function $\bar{\beta}_\phi(\eta, \tau, \sigma^2)$ be defined by $\bar{\beta}_\phi(Q_d R\mu, L_d\mu, \sigma^2) = \beta_\phi(\mu, \sigma^2)$ and let $\beta_\phi^i(\tau, \sigma^2)$ (resp., $\beta_\phi^{ij}(\tau, \sigma^2)$) be the derivative of $\bar{\beta}_\phi(\eta, \tau, \sigma^2)$ with respect to the ith (resp., ith and jth) component of η, evaluated at $\eta = 0$ (these derivatives always exist) A test ϕ in $H_d(\alpha)$ is said to be locally (near H_0) strictly unbiased if

(a) $\phi \in H_d^*(\alpha)$,

(b) $\beta_\phi^i(\tau, \sigma^2) = 0$ for all i, τ, and σ^2,

(c) the matrix $B_\phi(\tau, \sigma^2) = \| \beta_\phi^{ij}(\tau, \sigma^2) \|$ is positive definite for all τ and σ^2.

Clearly, (c) can be satisfied only if $d \in \Delta'$. Suppose then that $d \in \Delta'$ and that Q_d = identity (we have mentioned the fact that we can make this choice of Q_d when $d \in \Delta'$). For any ϕ satisfying (a), (b), (c) just above, det $B_\phi(\tau, \sigma^2)$ is the Gaussian curvature of the surface given by the graph of $\bar{\beta}_\phi(\eta, \tau, \sigma^2)$ as a function of η for fixed τ, σ^2, at $\eta = 0$. A test ϕ is defined in [11] to be of type D if it maximizes this curvature for all τ and σ^2, among all locally strictly unbiased tests. This criterion of optimality, although a local one, has certain appealing features; for example, it is invariant under all one-to-one transformations of the parameter space which leave $\eta = 0$ fixed and which at $\eta = 0$ are twice differentiable with non-vanishing Jacobian [11]. Now, since without loss of generality we are taking Q_d = identity, we can compare the behavior of the type D tests for various designs in Δ', assuming b_d to be the same for all d in Δ'. A design for which the Gaussian curvature at $\eta = 0$ of the test of maximum Gaussian curvature (for a given design) is a maximum (over all designs) is thus, if it exists, that d which maximizes $\max_{\phi_d} \det B_{\phi_d}(\tau, \sigma^2)$ simultaneously for all τ, σ^2. That such a design is precisely one which is D-optimum follows immediately from the following lemma[4] (there is an obvious analogue when σ^2 is known):

LEMMA 2.1. *For d in Δ' and $0 < \alpha < 1$, the test $F_{d,\alpha}$ is of type D.*

PROOF. $F_{d,\alpha}$ is clearly locally strictly unbiased. We again put Q_d = identity, and a nonsingular linear transformation reduces the proof to the case where G_d = identity (see Section 2A(c)), so that $\nu = \eta$. Wald's theorem can then be stated as

$$\int_{\eta'\eta=c} [\bar{\beta}_{F_{d,\alpha}}(\eta, \tau, \sigma^2) - \alpha] A(d\eta) \geq \int_{\eta'\eta=c} [\bar{\beta}_\phi(\eta, \tau, \sigma^2) - \alpha] A(d\eta) \tag{2.5}$$

for every $c > 0$, $\sigma^2 > 0$, and ϕ in $H_d^*(\alpha)$, where $A(d\eta)$ is Lebesgue measure on the sphere $\eta'\eta = c$. Noting that

$$\int_{\eta'\eta=c} \eta_i \eta_j A(d\eta) = \begin{cases} K(c, r) & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \tag{2.6}$$

where η_i is the ith component of η and $K(c, r)$ is positive and depends only on c

[4] The author understands that Isaacson gave a longer, unpublished proof, earlier.

and r, we obtain from (2.5) by normalizing properly and letting $c \to 0$, for any ϕ satisfying conditions (a) and (b) above,

$$\sum_{i=1}^{r} \beta_{F_{d,\alpha}}^{ii}(\tau, \sigma^2) \geqq \sum_{i=1}^{r} \beta_{\phi}^{ii}(\tau, \sigma^2). \tag{2.7}$$

Since $B_{F_{d,\alpha}}(\tau, \sigma^2)$ is a constant times the identity in our reduction, using the inequality of the geometric and arithmetic means and the fact that the determinant of a positive-definite matrix is no greater than the product of its diagonal elements, we obtain (omitting some appearances of τ, σ^2),

$$\begin{aligned} \det B_\phi(\tau, \sigma^2) &\leqq \prod_{i=1}^{r} \beta_\phi^{ii}(\tau, \sigma^2) \leqq \left[\sum_{i=1}^{r} \beta_\phi^{ii}/r\right]^r \\ &\leqq \left[\sum_{i=1}^{r} \beta_{F_{d,\alpha}}^{ii}/r\right]^r = \prod_{i=1}^{r} \beta_{F_{d,\alpha}}^{ii} = \det B_{F_{d,\alpha}}(\tau, \sigma^2), \end{aligned} \tag{2.8}$$

which completes the proof.

To summarize, D-optimality and L-optimality, although local properties, seem more reasonable criteria than E-optimality, which is tied to the ad hoc assumption that $F_{d,\alpha}$ should always be used; M-optimality (and to a lesser extent L-optimality) seems difficult to verify in many examples.

2C. *Relationship among optimality criteria in symmetric cases.* For future reference we state the following simple result (which was alluded to in Section 0 in reference to the relation between [1] and [2]):

LEMMA 2.2. *Suppose b_d is constant for d in Δ'. If d^* is D-optimum and $\bar{V}_{d^*}$ is a multiple of the identity, then d^* is E-optimum and L-optimum.*

PROOF. E-optimality is obvious from the nature of $\bar{V}_{d^*}$. If d^* were not L-optimum, since $F_{d^*,\alpha}$ has property 2A(c), for some other design d' there would by (2.2) be an associated test $\phi_{d'}$ in $H_{d'}^*(\alpha)$ with

$$\inf_{\tau,\sigma^2} \det B_{\phi_{d'}}(\tau, \sigma^2) > \det B_{F_{d^*,\alpha}}(\tau, \sigma^2) \tag{2.9}$$

(the right side of (2.9) is constant); by Lemma 2.1, equation (2.9) is a fortiori true if $\phi_{d'}$ is replaced by $F_{d',\alpha}$; this yields the contradiction that $\det \bar{V}_{d'} < \det \bar{V}_{d^*}$.

In many examples of Case I where symmetrical designs exist, the condition on $\bar{V}_{d^*}$ in the hypothesis of Lemma 2.2 will be obvious. In Case II, as discussed in Section 3A, it is often convenient to write the normal equations in the form $C_d t_d^* = Z_d$, where C_d is a $u \times u$ matrix of rank $u - 1$, Z_d is a u-vector of linear forms in Y with covariance matrix C_d, and for any solution t_d^* of these equations one obtains the best linear estimator of any contrast $\sum_1^u c_i \mu_i$ with $\sum c_i = 0$ by forming $\sum c_i t_{di}^*$ where the t_{di}^* are the components of t_d^*. Clearly, $P t_d^*$ is the b.l.e. t_d of $R_{II}\,\mu$. Hence, if every diagonal element of C_d has the same value and if all off-diagonal elements have the same value, the fact that the first row of the orthogonal matrix $\bar{O}$ defined in Case II of Section 1 is constant immediately yields the fact (see Section 3A) that $\bar{V}_d^{-1} = P C_d P'$ is a multiple of the identity, so that

Lemma 2.2 may be applicable in such cases. For future reference, we state this simple computation (put $a + (u - 1)c = 0$) in

LEMMA 2.3. *If U is a $u \times u$ matrix with diagonal elements a and off-diagonal elements c, then*

$$\bar{O}U\bar{O}' = \begin{pmatrix} a + (u - 1)c & 0 \\ 0 & (a - c)I_{u-1} \end{pmatrix}. \tag{2.10}$$

We remark that the form of R_{II} (associated with $\bar{O}$) used here makes computations and proofs simpler and emphasizes more the role of symmetry (e.g., as it appears in the form of $\bar{V}_d^{-1}$ just noted, when C_d has appropriate symmetry), than would be the case if R_{II} were replaced by a matrix obtained by adjoining a column of 1's and $m - u$ columns of 0's to I_{u-1}, as in [1] and [2].

3. Optimality of symmetrical designs.

3A. *Preliminaries.* The results of this section will be proved for the case where σ^2 is unknown, the other case being handled similarly. The setting of two-way heterogeneity where the LS design is employed is much easier to analyze (and thereby obtain an optimality proof) than other settings considered in Section 3B such as those where the YS and BIBD are used (and the remarks at the end of Section 2 indicate how this analysis can be made even simpler than in [1] and [2]). The reason for this is that in this setting where the LS is used, whether μ is considered to have $3u$ components (u each for row, column, and treatment effects in the $u \times u$ case) or $3u - 2$ components (to make $X_d'X_d$ nonsingular when $s_d = b_d = u - 1$), $X_d'X_d$ becomes particularly simple, having large blocks of 1's (each row and column occur together once, etc.) or multiples of an identity (rows by rows, etc.) in the former case, and large blocks of 0's (especially if $\bar{O}$ is used in reducing X_d) and multiples of an identity, in the latter. Other design situations yield more complicated forms of $X_d'X_d$. Therefore, although the examples of Section 3C could be analyzed in a manner analogous to that used for the LS in [1] and [2], it appears algebraically simpler to use the incomplete block design analysis of Bose [4], to which end we now briefly outline the notation. Of course, we are concerned here with the more difficult Case II, which includes most of the important examples.

The form of the Z_d and C_d mentioned in Section 2C depends on the design setting and, in particular, in this section, on whether we are in a setting of one-way or two-way heterogeneity of (for example) soil (since all block sizes will be the same in our example of the former, it could be considered as a special case of the latter under further restrictions on μ). We shall first state the pertinent results which apply in both of these settings, and then specify the particular forms (see [4] for details). The $u \times u$ symmetric matrix C_d has row (or column) sums equal to zero, and the sum of the components of the u-vector Z_d is zero The covariance matrix of Z_d is $\sigma^2 C_d$ and the expected value of Z_d is $C_d\mu^{(u)}$, where $\mu^{(u)}$ is the vector of the first u components of μ. We may assume $d \varepsilon \Delta'$, which means the design d is connected and that C_d has rank $u - 1$. If t_d^* satisfies $C_dt_d^* =$

Z_d and P is the $(u-1) \times u$ matrix defined in Case II in Section 1, then $t_d = Pt_d^*$ is the vector if b.l.e.'s of $R_{II}\mu$; the last $u-1$ rows of the equation $\bar{O}C_d\bar{O}'\bar{O}t_d^* = \bar{O}Z_d$ are thus $PC_dP't_d = PZ_d$ (the first row and column of $\bar{O}C_d\bar{O}'$ are zero), so that $t_d = (PC_dP')^{-1}PZ_d$ (the inverse may be taken for d in Δ') and thus the components of t_d have covariance matrix $(PC_dP')^{-1}$.

In the one-way heterogeneity setting we have u treatments, to be planted in b blocks; in our example, each block will contain the same number k of plots, one "planting" to be allowed per plot. The component of Y corresponding to an appearance of treatment i in block j has expected value $\mu_i + b_j$; thus, $m = u + b$, with $\mu_{u+j} = b_j$. Let n_{dij} be the number of appearances of treatment i in block j. We do *not* restrict n_{dij} to be 0 or 1, as is often done. Thus, D consists of those d for which X_d is any matrix of 0's and 1's for which each row contains exactly one 1 among the first m elements and one 1 among the last b elements and for which the last b columns each contain k one's; of course, $N = bk$. Let $r_{di} = \sum_j n_{dij}$ = number of replications of treatment i, let T_{di} = sum of all components of Y corresponding to treatment i, and let B_{dj} = sum of all components of Y arising from block j. The ith component Z_{di} of Z_d ("adjusted yield of treatment i") is $Z_{di} = T_i - \sum_j n_{ij} B_j/k$, and the (i, j)th component c_{dij} of C_d is

$$c_{dij} = \delta_{ij} r_{di} - \lambda_{dij}/k, \tag{3.1}$$

where δ_{ij} is the Kronecker delta and $\lambda_{dij} = \sum_s n_{dis} n_{djs}$.

In the setting of two-way heterogeneity, we have u treatments and a $k_1 \times k_2$ array of plots, and the expected value of a component of Y corresponding to treatment i in row j and column h is $\mu_i + b_j^{(1)} + b_h^{(2)}$; thus, $m = u + k_1 + k_2$ with $b_j^{(1)} = \mu_{m+j}$ and $b_h^{(2)} = \mu_{m+k_1+h}$. Let $n_{dij}^{(1)}$ (resp., $n_{dih}^{(2)}$) be the number of times treatment i appears in row j (resp., column h), and let T_{di} be as before and $B_{dj}^{(1)}$ (resp., $B_{dh}^{(2)}$) be the sum corresponding to the jth row (resp., hth column). r_{di} is as above, while $\lambda_{dij}^{(q)} = \sum_s n_{dis}^{(q)} n_{djs}^{(q)}$ for $q = 1, 2$. In this case $Z_{di} = T_{di} - \sum_j n_{dij}^{(1)} B_{dj}^{(1)}/k_2 - \sum_h n_{dih}^{(2)} B_{dh}^{(2)}/k_1 + r_{di} \sum_s T_{ds}/k_1 k_2$ and

$$c_{dij} = \delta_{ij} r_{di} - \frac{\lambda_{dij}^{(1)}}{k_2} - \frac{\lambda_{dij}^{(2)}}{k_1} + \frac{r_{di} r_{dj}}{k_1 k_2}. \tag{3.2}$$

Many other design settings can be treated similarly; the above two will be used in the examples of Section 3C to illustrate our methods of proving optimality.

3B. *Algebraic results.* We now demonstrate the algebraic results used in proving optimality in the examples of Section 3C and which will be useful in other examples of Case II. The results proved here are meant to apply elsewhere than in the settings of Section 3A. We suppose in the present Section 3B that we are given a class $\{K_d, d \,\varepsilon\, \Delta'\}$ of $u \times u$ symmetric nonnegative definite matrices of rank $u-1$ with row and column sums zero and define $W_d = PK_dP'$ (in our applications, $W_d = \bar{V}_d^{-1}$). The elements of $\bar{O}$, K_d, and W_d will be denoted by $\bar{o}_{ij}$,

$k_{d_{ij}}$, and $w_{d_{ij}}$, respectively. In Lemma 3.2 we consider an orthogonal matrix $\tilde{O} = \| o_{ij} \|$, not necessarily $\bar{O}$, and a diagonal matrix $D = \| d_{ij} \|$.

Our first lemma merely translates into terms of K_d the obvious fact that, if W_{d^*} has equal eigenvalues and if the sum of the eigenvalues (= trace) of W_d is a maximum for $d = d^*$, then the product of eigenvalues (= determinant) of W_d is a maximum for $d = d^*$.

LEMMA 3.1. *If all diagonal elements of K_{d^*} are equal and all off-diagonal elements of K_{d^*} are equal and $\sum_i k_{d_{ii}}$ is a maximum for $d = d^*$, then* det W_d *is a maximum for $d = d^*$.*

PROOF. Since $\bar{o}_{ij} = 1/\sqrt{u}$ and $\sum_{i,j} k_{d_{ij}} = 0$, the upper left-hand element of $\bar{O}K_d\bar{O}'$ is zero. Since the traces of $\bar{O}K_d\bar{O}'$ and K_d are equal, we conclude that the traces of K_d and W_d are equal, so that the trace of W_d is a maximum for $d = d^*$. The result now follows from Lemma 2.3 (follow the steps of (2.8) with W_d for B_ϕ and W_{d^*} for $B_{P_{d,\alpha}}$).

We shall actually prove in Theorems 3.1 and 3.2 that the trace of the matrix PC_dP' is a maximum and that all eigenvalues are equal when d is a BBD or GYS, so that Lemma 3.1 is relevant. However, there are settings in which the next three lemmas are more useful for proving D- or E-optimality directly when the hypothesis of Lemma 3.1 is difficult to verify or is false.

LEMMA 3.2. *For $u > 1$ if $\tilde{O}$ is orthogonal $u \times u$, D is diagonal $u \times u$, K is symmetric nonnegative definite $u \times u$ with row and column sums zero, and $\tilde{O}D\tilde{O}' = K$, then*

$$(3.3) \qquad \left(\frac{u-1}{u}\right)^u \left(\prod_{i=1}^{u-1} d_{ii}\right)^{u/(u-1)} \leqq \prod_{i=1}^{u} k_{ii}.$$

PROOF. We assume $d_{uu} = 0 < d_{ii}$ for $i < u$, or the result is trivial. Since, then,

$$(3.4) \qquad 0 = \sum_{i=1}^{u}\sum_{j=1}^{u} k_{ij} = \sum_{i=1}^{u}\sum_{j=1}^{u}\sum_{s=1}^{u-1} o_{is}o_{js}d_{ss} = \sum_{s=1}^{u-1} d_{ss}\left(\sum_{i=1}^{u} o_{is}\right)^2,$$

we conclude that the first $u - 1$ columns of $\tilde{O}$ are orthogonal to the vector of ones. Hence, $o_{ju} = 1/\sqrt{u}$ (or its negative, which is treated in the same way).

Let the coordinates of a point ϵ in $u(u-1)$-dimensional Euclidean space be denoted by ϵ_{ij} $(i = 1, \cdots, u; j = 1, \cdots, u-1)$, and let B be the set of points ϵ in this space for which all $\epsilon_{ij} \geqq 0$, for which $\sum_j \epsilon_{ij} = (u-1)/u$ for all i, and for which $\sum_i \epsilon_{ij} = 1$ for all j. We shall prove below that ϵ in B implies

$$(3.5) \qquad \prod_{i=1}^{u}\left(\sum_{j=1}^{u-1} \epsilon_{ij}d_{jj}\right) \geqq \left(\frac{u-1}{u}\right)^u \left(\prod_{i=1}^{u-1} d_{ii}\right)^{u/u-1};$$

since the left side of (3.5) with $\epsilon_{ij} = o_{ij}^2$ gives the right side of (3.3) and since the restrictions on the ϵ_{ij} in B must be satisfied by the o_{ij}^2 (the orthogonality restrictions on the o_{ij} are omitted in defining B), (3.5) implies (3.3).

Call the left side of (3.5) $f(\epsilon)$. It is easy to verify that $-\log f(\epsilon)$ is convex in ϵ on $u(u-1)$-space, and hence on B. Moreover, B is a convex body in $u(u-1)$

space, and any extreme point of B is either

$$(3.6)\qquad \begin{pmatrix} \epsilon_{11} & \cdots & \epsilon_{1,u-1} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \epsilon_{u1} & \cdots & \epsilon_{u,u-1} \end{pmatrix} = \begin{pmatrix} \frac{u-1}{u} I_{u-1} \\ \frac{1}{u} \cdots\cdots\cdots \frac{1}{u} \end{pmatrix}$$

or is obtained by permuting the rows of the matrix on the right side of (3.6). Since a convex function on a convex set attains its maximum at an extreme point, we conclude that the minimum of f is attained at one of these extreme points. But f has the same value at any of these extreme points, namely,

$$(3.7)\qquad \min_B f(\epsilon) = \left(\sum_{i=1}^{u-1} d_{ii}/u\right) \prod_{i=1}^{u-1} \left(\frac{u-1}{u} d_{ii}\right).$$

Thus, it remains only to prove that the right side of (3.7) is no less than the right side of (3.5), i.e., that

$$(3.8)\qquad \prod_{i=1}^{u-1} d_{ii}^{1/(u-1)} \leqq \sum_{i=1}^{u-1} d_{ii}/(u-1);$$

but (3.8) is merely the well-known inequality between the geometric and arithmetic means.

The form of Lemma 3.2 which is useful in many applications is the following:

LEMMA 3.3. *If* $\prod_1^u k_{dii}$ *is a maximum for* $d = d^*$ *and if* K_{d^*} *has all diagonal elements equal and all off-diagonal elements equal, then* det W_d *is a maximum for* $d = d^*$.

PROOF: We use Lemma 3.2 with the product on the left side of (3.3) going from 2 to u, in order to conform to previous notation. In this form, with $\tilde{O} = \bar{O}$, it follows from Lemma 2.3 that the left and right sides of (3.3) are equal for $K = K_{d^*}$. Hence, from Lemma 3.2, $\prod_i w_{dii}$ is a maximum for $d = d^*$. Since $\prod_i w_{dii} \geqq$ det W_d with equality for the diagonal matrix W_{d^*}, the proof is complete.

The following lemma could be used in the case of the YS, and in more complicated problems where D-optimality is hard to prove or false, to prove E-optimality directly (i.e., without the use of Lemma 2.2):

LEMMA 3.4. *For* $u > 1$, *if* $m(W_d)$ *is the minimum eigenvalue of* W_d, *then*

$$(3.9)\qquad m(W_d) \leqq \frac{u}{u-1} \min_i k_{dii};$$

if all diagonal elements of K_d *are equal and all off-diagonal elements are equal, equality holds in* (3.9).

PROOF. Let δ_i be a u-vector with ith element one and all other elements zero. Let $\xi_i = P\delta_i$. Clearly, $\sqrt{u/(u-1)}\, \xi_i$ has unit length. Hence,

$$(3.10)\qquad \begin{aligned} k_{dii} &= \delta_i' K_d \delta_i = (\bar{O}\delta_i)'(\bar{O} K_d \bar{O}')(\bar{O}\delta_i) \\ &= \xi_i' W_d \xi_i \geqq \frac{u-1}{u} \min_{a'a=1} a' W_d a = \frac{u-1}{u} m(W_d), \end{aligned}$$

which proves (3.9); the result on equality follows from Lemma 2.3.

The results for Case I analogous to those proved for Case II in this subsection are trivial (since in Case I the analogue of K_d will be nonsingular and K_{d^*} will be a multiple of the identity), and will be omitted.

3C. *Examples.*[5] (1). *Optimality of* BIBD's. In the setting of one-way heterogeneity described in Section 3A (with $u > 1$), suppose b, u, and k to be such that there exists a design d^* for which all n_{d^*ij} are k/u if k/u is an integer and are either of the two integers closest to k/u otherwise, for which all r_{d^*i} are equal, and for which all λ_{d^*ij} are equal for $i \neq j$. Such a design is called a BIBD if $k < u$, but we do not impose this last restriction here, and therefore call such a design a Balanced Block Design (BBD). (For example, if $b = 2$, $u = 2$, $k = 3$, such a d^* is that for which $n_{d^*11} = n_{d^*22} = 1$ and $n_{d^*12} = n_{d^*21} = 2$.) Our result is:

THEOREM 3.1. *If a* BBD d^* *exists, it is D-optimum, E-optimum, and L-optimum*

PROOF. From (3.1) we have

$$\sum_{i=1}^{u} c_{dii} = N - \sum_i \sum_j n_{dij}^2/k; \tag{3.11}$$

since $\sum_i \sum_j n_{dij} = N$, it is clear that (3.11) is a maximum for $d = d^*$. The result now follows from Lemma 3.1 and Lemma 2.2.

(2). *Optimality of* YS's. In the setting of two-way heterogeneity described in Section 3A (with $u > 1$), suppose k_1, k_2, and u to be such that there exists a design d^* for which all r_{d^*i} are equal, for which all $\lambda_{d^*ij}^{(1)}$ are equal for $i \neq j$, for which all $\lambda_{d^*ij}^{(2)}$ are equal for $i \neq j$, and for which all $n_{d^*ij}^{(q)}$ are equal to k_q/u if k_q/u *is an integer and are either of the two integers closest to* k_q/u *otherwise* ($q = 1, 2$). Thus, d^* is a BBD when either the rows or the columns are considered to be the blocks. Such a design d^* is usually called a YS if $k_1 < u$ (and k_2/u is an integer); we do not impose this condition, and shall hence call such a design d^* a Generalized Youden Square (GYS). (For example, if $u = 2$, $k_1 = 4$, $k_2 = 3$, such a design d^* is easily constructed.) If $k_1 = k_2 = u$, such a d^* is of course a LS. Our result is:

THEOREM 3.2. *If k_1/u or k_2/u is an integer and if a* GYS d^* *exists, then* d^* *is D-optimum, E-optimum, and L-optimum.*

PROOF. We shall show that $\sum_i c_{dii}$ is a maximum for $d = d^*$; Lemma 3.1 then yields the desired result. In this proof only we write $[x]$ = greatest integer $\leq x$. Let r be an integer. Subject to the restrictions that $\sum_1^k m_j = r$ and that all m_j are integers, the expression $\sum_1^k m_j^2$ is minimized by taking $k - r + k[r/k]$ of the m_j to be $[r/k]$ and $r - k[r/k]$ of them to be $[r/k] + 1$, the corresponding minimum of $\sum m_j^2$ being $r + (2r - k)[r/k] - k[r/k]^2 = h(r, k)$ (say). We may assume

[5] The Editor has informed the author that E-optimality of the BIBD's (as a subclass of the BBD's) has been proved independently by V. L. Mote, and that the minimization of the average variance (see numbered paragraph 2 of Section 5) and of the generalized variance (i.e., the attainment of D-optimality) achieved by the BIBD's and YS's (a subclass of the GYS's) has been proved independently by A. M. Kshirsagar; both of these authors prove their results under the restriction and that the n_{dij} and $n_{dij}^{(q)}$ are all 0 or 1. Under this restriction, these special cases of the results of this paper are a consequence of the following line of argument: the trace of C_d is the same for all d, and the results follow at once from the symmetry of the BIBD and YS.

space, and any extreme point of B is either

$$(3.6)\qquad \begin{pmatrix} \epsilon_{11} & \cdots & \epsilon_{1,u-1} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \epsilon_{u1} & \cdots & \epsilon_{u,u-1} \end{pmatrix} = \begin{pmatrix} \frac{u-1}{u} I_{u-1} \\ \frac{1}{u} \cdots\cdots\cdots \frac{1}{u} \end{pmatrix}$$

or is obtained by permuting the rows of the matrix on the right side of (3.6). Since a convex function on a convex set attains its maximum at an extreme point, we conclude that the minimum of f is attained at one of these extreme points. But f has the same value at any of these extreme points, namely,

$$(3.7)\qquad \min_B f(\epsilon) = \left(\sum_{i=1}^{u-1} d_{ii}/u\right) \prod_{i=1}^{u-1} \left(\frac{u-1}{u} d_{ii}\right).$$

Thus, it remains only to prove that the right side of (3.7) is no less than the right side of (3.5), i.e., that

$$(3.8)\qquad \prod_{i=1}^{u-1} d_{ii}^{1/(u-1)} \leqq \sum_{i=1}^{u-1} d_{ii}/(u-1);$$

but (3.8) is merely the well-known inequality between the geometric and arithmetic means.

The form of Lemma 3.2 which is useful in many applications is the following:

LEMMA 3.3. *If* $\prod_1^u k_{dii}$ *is a maximum for* $d = d^*$ *and if* K_{d^*} *has all diagonal elements equal and all off-diagonal elements equal, then* det W_d *is a maximum for* $d = d^*$.

PROOF: We use Lemma 3.2 with the product on the left side of (3.3) going from 2 to u, in order to conform to previous notation. In this form, with $\tilde{O} = \bar{O}$, it follows from Lemma 2.3 that the left and right sides of (3.3) are equal for $K = K_{d^*}$. Hence, from Lemma 3.2, $\prod_i w_{dii}$ is a maximum for $d = d^*$. Since $\prod_i w_{dii} \geqq$ det W_d with equality for the diagonal matrix W_{d^*}, the proof is complete.

The following lemma could be used in the case of the YS, and in more complicated problems where D-optimality is hard to prove or false, to prove E-optimality directly (i.e., without the use of Lemma 2.2):

LEMMA 3.4. *For* $u > 1$, *if* $m(W_d)$ *is the minimum eigenvalue of* W_d, *then*

$$(3.9)\qquad m(W_d) \leqq \frac{u}{u-1} \min_i k_{dii};$$

if all diagonal elements of K_d *are equal and all off-diagonal elements are equal, equality holds in* (3.9).

PROOF. Let δ_i be a u-vector with ith element one and all other elements zero. Let $\xi_i = P\delta_i$. Clearly, $\sqrt{u/(u-1)}\,\xi_i$ has unit length. Hence,

$$(3.10)\qquad \begin{aligned} k_{dii} &= \delta_i' K_d \delta_i = (\bar{O}\delta_i)'(\bar{O}K_d\bar{O}')(\bar{O}\delta_i) \\ &= \xi_i' W_d \xi_i \geqq \frac{u-1}{u} \min_{a'a=1} a' W_d a = \frac{u-1}{u} m(W_d), \end{aligned}$$

which proves (3.9); the result on equality follows from Lemma 2.3.

The results for Case I analogous to those proved for Case II in this subsection are trivial (since in Case I the analogue of K_d will be nonsingular and K_{d^*} will be a multiple of the identity), and will be omitted.

3C. *Examples.*[5] (1). *Optimality of* BIBD's. In the setting of one-way heterogeneity described in Section 3A (with $u > 1$), suppose b, u, and k to be such that there exists a design d^* for which all n_{d^*ij} are k/u if k/u is an integer and are either of the two integers closest to k/u otherwise, for which all r_{d^*i} are equal, and for which all λ_{d^*ij} are equal for $i \neq j$. Such a design is called a BIBD if $k < u$, but we do not impose this last restriction here, and therefore call such a design a Balanced Block Design (BBD). (For example, if $b = 2$, $u = 2$, $k = 3$, such a d^* is that for which $n_{d^*11} = n_{d^*22} = 1$ and $n_{d^*12} = n_{d^*21} = 2$.) Our result is:

THEOREM 3.1. *If a* BBD d^* *exists, it is D-optimum, E-optimum, and L-optimum*

PROOF. From (3.1) we have

$$\sum_{i=1}^{u} c_{dii} = N - \sum_i \sum_j n_{dij}^2/k; \tag{3.11}$$

since $\sum_i \sum_j n_{dij} = N$, it is clear that (3.11) is a maximum for $d = d^*$. The result now follows from Lemma 3.1 and Lemma 2.2.

(2). *Optimality of* YS's. In the setting of two-way heterogeneity described in Section 3A (with $u > 1$), suppose k_1, k_2, and u to be such that there exists a design d^* for which all r_{d^*i} are equal, for which all $\lambda_{d^*ij}^{(1)}$ are equal for $i \neq j$, for which all $\lambda_{d^*ij}^{(2)}$ are equal for $i \neq j$, and for which all $n_{d^*ij}^{(q)}$ are equal to k_q/u if k_q/u is an integer and are either of the two integers closest to k_q/u otherwise ($q = 1, 2$). Thus, d^* is a BBD when either the rows or the columns are considered to be the blocks. Such a design d^* is usually called a YS if $k_1 < u$ (and k_2/u is an integer); we do not impose this condition, and shall hence call such a design d^* a Generalized Youden Square (GYS). (For example, if $u = 2$, $k_1 = 4$, $k_2 = 3$, such a design d^* is easily constructed.) If $k_1 = k_2 = u$, such a d^* is of course a LS. Our result is:

THEOREM 3.2. *If k_1/u or k_2/u is an integer and if a* GYS d^* *exists, then* d^* *is D-optimum, E-optimum, and L-optimum.*

PROOF. We shall show that $\sum_i c_{dii}$ is a maximum for $d = d^*$; Lemma 3.1 then yields the desired result. In this proof only we write $[x]$ = greatest integer $\leqq x$. Let r be an integer. Subject to the restrictions that $\sum_1^k m_j = r$ and that all m_j are integers, the expression $\sum_1^k m_j^2$ is minimized by taking $k - r + k[r/k]$ of the m_i to be $[r/k]$ and $r - k[r/k]$ of them to be $[r/k] + 1$, the corresponding minimum of $\sum m_j^2$ being $r + (2r - k)[r/k] - k[r/k]^2 = h(r, k)$ (say). We may assume

[5] The Editor has informed the author that E-optimality of the BIBD's (as a subclass of the BBD's) has been proved independently by V. L. Mote, and that the minimization of the average variance (see numbered paragraph 2 of Section 5) and of the generalized variance (i.e., the attainment of D-optimality) achieved by the BIBD's and YS's (a subclass of the GYS's) has been proved independently by A. M. Kshirsagar; both of these authors prove their results under the restriction and that the n_{dij} and $n_{dij}^{(q)}$ are all 0 or 1. Under this restriction, these special cases of the results of this paper are a consequence of the following line of argument: the trace of C_d is the same for all d, and the results follow at once from the symmetry of the BIBD and YS.

k_1/u is an integer. From (3.2) we have, for any d,

$$(3.12)\quad k_1k_2(k_1k_2 - \sum_i c_{dii}) \geqq \sum_i \{k_2h(r_{di}, k_2) - r_{di}^2\} + \sum_i k_1h(r_{di}, k_1),$$

with equality in the case of a GYS. The theorem will be proved if we show that each of the two sums on the right side of (3.12) attains its minimum for $d = d^*$. Now, $h(r, k) \geqq r^2/k$, since the latter is the minimum of $\sum m_j^2$ subject to $\sum m_j = r$ without the restriction that the m_j be integers. Hence, the first sum on the right side of (3.12) is at least zero. Moreover, this lower bound is achieved by the first sum on the right in (3.12) when $d = d^*$, since $r_{d^*i}/k_2 = k_1/u$ is an integer. It remains to consider the last sum of (3.12). We shall show that, subject to $\sum_1^m z_i = c$, the expression

$$(3.13)\qquad q(z_1, \cdots, z_m) = \sum_{i=1}^{m} \{(2z_1 - 1)[z_i] - [z_i]^2\}$$

is a minimum when all z_i are equal; putting $z_i = [r_{di}/k_1]$, we see that this will yield the desired conclusion regarding the last sum of (3.12). The proof regarding (3.13) is by induction: assuming the conclusion to be true of $m = M$, in proving the case $m = M + 1$ we may put $z_1 = \cdots = z_M = s$ and $z_{M+1} = c - Ms$ in (3.13). The resulting expression is continuous in s and, except on a discrete set, has a derivative with respect to s which is equal to $2M([s] - [c - Ms])$. The latter is $\leqq 0$ if $s < c/(M + 1)$ and is $\geqq 0$ if $s > c/(M + 1)$, so that $s = c/(M + 1)$ yields a minimum. This completes the proof of Theorem 3.2.

We remark that, without the assumption that k_1/u or k_2/u is an integer, the above proof fails and Lemma 3.3 also fails to be applicable generally. To see this, consider the case $k_1 = k_2 = 6$, $u = 4$. A GYS d^* exists here, e.g., that one whose successive rows are (134324), (412233), (241342), (124123), (313412), (321441). We obtain $c_{d^*ii} = 25/4$ for all i. Let d' be the design whose rows are (133442), (213344), (421334), (442133), (344213), and (334421). Then $c_{d'11} = c_{d'22} = 5$, $c_{d'33} = c_{d'44} = 8$, $c_{d'12} = -1$, $c_{d'34} = -4$, and all other $c_{d'ij} = -2$. Thus, we obtain $\sum_i c_{d'ii} = 26 > 25 = \sum_i c_{d^*ii}$ and even $\prod_i c_{d'ii} = 1600 > (25/4)^4 = \prod_i c_{d^*ii}$. However, det $\bar{V}_{d'}^{-1} = 576 < (25/3)^3 =$ det $\bar{V}_{d^*}^{-1}$. Thus, between the designs d^* and d', the former is D-optimum, although Lemmas 3.1 and 3.3 cannot be used to prove it. Lemma 3.4 could still have been used to prove the E-optimality of d^* directly.

(3) *Other examples.* Many other design settings can be analyzed in a manner differing only slightly from the above examples and we mention but a few. One can treat similarly problems where the test concerns the b_j and $b_j^{(q)}$ of Section 3A. Problems involving Graeco-Latin Squares or higher Latin Squares, with or without replications, admit similar treatments. Higher-dimensional analogues (more than two directions of heterogeneity) can also be considered in a like fashion, as can complete or partial factorial arrangements. Many of the Case I analogues, such as the analogue of the BIBD treatment which assumes the b_j to be known, are trivial.

Other problems such as those for which E-optimality is considered in [2] (e.g., Hotelling's weighing problem and certain problems in the analysis of covariance) could be considered regarding D- and L-optimality by similar methods.

The treatment of some problems is in part parallel, but entails other considerations in addition to symmetry; such a problem is to test whether a regression function $\sum_{i=1}^{r} \mu_i f_i(x)$ is actually such that $\mu_1 = \cdots = \mu_r = 0$, where the f_i are given and N x's must be chosen from a given region of some space. (Many problems in the analysis of covariance involve similar considerations.) D- and E-optimality are also relevant in estimation problems (see Section 5.2).

The consideration of some of these other examples will appear elsewhere, in a paper by J. Wolfowitz and the author.

4. Nonoptimality of symmetrical nonrandomized designs among randomized designs.[1]

4.A. Case I. We consider here the simplest general setting of Case I, namely, the extension of the example of Section 1 to more observations N and more treatments u. Other examples, such as the Case I analogues of the examples of Section 4B, have parallel analyses, and we omit them. We shall carry out the treatment when σ^2 is unknown, the treatment when σ^2 is known being similar. The underlying probabilistic property (of the normal distribution) which is relevant here will now be stated in a lemma. Let U/σ^2 have a non-central χ^2 distribution with N_1 degrees of freedom (d.f.) and non-central parameter $\lambda = EU/\sigma^2 - N_1$, and let V/σ^2 have the central χ^2 distribution with N_2 d.f., with U and V independent. Let $P_{N_1,N_2}(\lambda, \alpha)$ denote the power function of the F-test of size α for testing $\lambda = 0$ based on N_2U/N_1V, and, as in (1.3), let $g_{N_1,N_2}(\alpha)$ denote the derivative of this power function with respect to λ at $\lambda = 0$.

Lemma 4.1. *If $N_1 \leqq N_1'$ and $N_1 + N_2 \geqq N_1' + N_2'$ with at least one of these a strict inequality, then $P_{N_1,N_2}(\lambda; \alpha) > P_{N_1',N_2'}(\lambda, \alpha)$ for $\lambda > 0$ and $0 <$ [...] and $g_{N_1,N_2}(\alpha) > g_{N_1',N_2'}(\alpha)$ for $0 < \alpha < 1$.*

Proof. Let U/σ^2 have a χ^2 distribution with parameter λ and N_1 d.f., V_1/σ^2, V_2/σ^2, and V_3/σ^2 have central χ^2 distributions with N_2', $N_1' - N$[...] $N_1 + N_2 - N_1' - N_2'$ d.f., respectively (if any of the d.f.'s is 0, so is the [corre]sponding V_i). U, V_1, V_2, V_3 are independent. For testing the hypothesis λ [...] against alternatives $\lambda > 0$ based on U, V_1, V_2, V_3, it is easy to prove that [...] F-test based on $N_2U/N_1(V_1 + V_2 + V_3)$ is UMP unbiased of size α and is [of] type A, and is the unique (up to sets of measure zero) test with each of these properties; in particular, this is true in comparison with the F-test based on $N_2'(U + V_2)/N_1'V_1$, which proves the lemma.

The above lemma indicates both that the numerator d.f. should be as small as possible without affecting λ, which is also true when σ^2 is known, and also that for fixed $N_1 + N_2$, decreasing N_1 helps even more if σ^2 is *unknown*, since N_2 is increased (compare (4.5) and (4.7) below).

We now consider the following problem. Y_{ij} are independent and normally distributed random variables with unknown mean μ_i ($j = 1, \cdots, n_i$, $i = 1$,

$\cdots, u)$ and variance σ^2 (we use a convenient notation for the example, rather than that introduced in Section 1). The problem is to test $H_0: \mu_1 = \mu_2 = \cdots = \mu_u = 0$, and a design d in Δ is a specification of nonnegative integers n_i whose sum is N. For any such d, we denote by $M(d)$ the set of i for which $n_i > 0$; by $k(d)$, the number of integers in $M(d)$; by τd, the design associated with the values $n_i = n^*_{\tau(i)}$ when d is associated with the values $n_i = n^*_i$, where τ is any element of the symmetric group S_u on u symbols; by δ_d, the design in Δ_R which assigns probability $1/u!$ to each τd for τ in S_u; by $f_{d,\alpha}$ the test associated with δ_d which is obtained by using the appropriate F-test of size α with whatever τd is chosen by δ_d. We shall also use the symbol $a_\phi(c)$ of (2.2), with $\psi(\mu) = \sum_1^u \mu_i^2$, and shall denote by a'_ϕ its derivative with respect to c at $c = 0$. We shall also use the symbols $g_{ij}(\alpha)$ introduced in (1.3). Our result, which implies that the "symmetrical" design associated with $k(d) = u$ and all n_i equal (or as nearly so as possible) is not L_α-optimum in Δ_R, and that the δ_d associated with the d for which $n_1 = N$ (this δ_d chooses each i with probability $1/u$ and takes all Y_{ij} with the chosen i) is locally best among the δ_d, is the following:

THEOREM 4.1. *For every d, α, and c,*

$$a_{F_{d,\alpha}}(c) \leqq a_{f_{d,\alpha}}(c); \tag{4.1}$$

$a'_{f_{d,\alpha}}$ is strictly decreasing in $k(d)$, and the same is true of $a_{f_{d,\alpha}}(c)$ for all c in some neighborhood of $c = 0$.

PROOF. (4.1) is trivial, and we proceed to the rest of the proof. The numerator $t'_d V_d^{-1} t_d$ of $F_{d,\alpha}$ is of course

$$U_d = \sum_{i \varepsilon M(d)} n_i \left(\sum_{j=1}^{n_i} Y_{ij}/n_i \right)^2,$$

and U_d/σ^2 has a χ^2 distribution with $k(d)$ d.f. and non-central parameter

$$\sum_{i \varepsilon M(d)} n_i \mu_i^2/\sigma^2.$$

The denominator of $F_{d,\alpha}$ has $N - k(d)$ d.f. Write $\lambda = \sum_1^u \mu_i^2/\sigma^2$. From (1.3) we have, as $\lambda \to 0$,

$$\begin{aligned} \beta_{f_{d,\alpha}}(\mu, \sigma^2) &= \sum_{\tau \varepsilon S_u} \beta_{F_{\tau d,\alpha}}(\mu, \sigma^2)/u! \\ &= \sum_{\tau \varepsilon S_u} [\alpha + g_{k(d),N-k(d)}(\alpha) \sum_{\tau \varepsilon S_u} n_{\tau(i)} \mu_i^2/\sigma^2 + 0(\lambda^2)]/u! \\ &= \alpha + g_{k(d),N-k(d)}(\alpha) \sum_i (\sum_\tau n_{\tau(i)}/u!) \mu_i^2/\sigma^2 + 0(\lambda^2) \\ &= \alpha + \frac{N}{u} g_{k(d),N-k(d)}(\alpha)\lambda + 0(\lambda^2). \end{aligned} \tag{4.2}$$

The desired conclusion now follows from Lemma 4.1.

Existing tables and charts of the power functions of the F-test and χ^2-test are presented in such forms (in terms of $\sqrt{\lambda/(k(d) + 1)}$, usually in inverted form and with wide spacing of arguments) as to make accurate comparisons of the

$\beta_{f_{d,\alpha}}$ difficult. This difficulty is made the worse by the fact that $\beta_{f_{d,\alpha}}$ is not (with an obvious exception) constant on the contour λ = constant, making it somewhat of a task to obtain $a_{f_{d,\alpha}}(c)$. It is not true, as might be supposed, that this minimum power on the contour λ = constant is always attained for a μ with all components equal, or else is always attained for a μ with all components except one equal to zero. To see this, consider the problem of Section 1 ($N = u = 2$, σ^2 known). Let C_α be the value such that, if Y is a normal random variable with 0 mean and unit variance, then $P\{ | Y | > C_\alpha\} = \alpha$. A direct computation of the power function of δ near $\lambda \equiv \mu_1^2 + \mu_2^2 = 0$ yields

$$\beta_\delta(\mu) = \alpha + \frac{C_\alpha \exp(-C_\alpha^2/2)}{2\sqrt{2\pi}} \cdot \{2(\mu_1^2 + \mu_2^2) + (C_\alpha^2 - 3)(\mu_1^4 + \mu_2^4)/3 + O(\lambda^3)\}. \tag{4.3}$$

Hence, when c is sufficiently small, the minimum of $\beta_\delta(\mu)$ on the contour $\lambda = c$, neglecting the term $O(\lambda^3)$, is located at $\mu_1 = \sqrt{c}$, $\mu_2 = 0$ (or $\mu_2 = \sqrt{c}$, $\mu_1 = 0$) if $C_\alpha \leqq \sqrt{3}$ and at $\mu_1 = \mu_2 = \sqrt{c/2}$ if $C_\alpha \geqq \sqrt{3}$. When we include terms of higher order in μ, it is no longer even evident that the minimum must be attained at one of these two values of μ.

We see from (4.3) that $g_{1,\infty}(\alpha) = (2\pi)^{-\frac{1}{2}}C_\alpha \exp(-C_\alpha^2/2)$ and it is not hard to show that $g_{2,\infty}(\alpha) = -\alpha(\log \alpha)/2$ (see [12], equation (6 27), where λ is our $\lambda/2$). Thus, a comparison of $a'_{f_{d,\alpha}}$ for $k(d) = 1$ and 2 is given in this example by the following table:

α	$g_{1,\infty}(\alpha)$	$g_{2,\infty}(\alpha)$
01	.037	023
.05	.114	075
10	.175	115
20	.225	.161
30	.242	.181
.50	.214	.173
.90	.050	.047

The following lemma shows that, as $\alpha \to 0$, the ratio of the second to third column above goes to 2 and, more generally, that $g_{i,\infty}(\alpha)/g_{j,\infty}(\alpha) \to j/i$ (this gives a comparison of the various δ_d for general N and u and for various $k(d)$ when σ^2 is known, as $\alpha \to 0$; see Lemma 4.3 for the case when σ^2 is unknown):

LEMMA 4.2. *As* $\alpha \to 0$,

$$g_{j,\infty}(\alpha) = -[1 + o(1)]\alpha(\log \alpha)/j. \tag{4.5}$$

PROOF. Fix j. Let k_α be such that if Y is a random variable with central χ^2 distribution with j d.f., then $P\{Y > k_\alpha\} = \alpha$ Let f_λ be the χ^2 density function with j d.f. and non-central parameter λ A simple calculation shows that $df_\lambda(u)/d\lambda$ at $\lambda = 0$ is $f_0(u)[(u/2j) - 1/2]$. Hence, as $k_\alpha \to \infty$,

$$g_{j,\infty}(\alpha) = \int_{k_\alpha}^{\infty} f_0(u)[(u/2j) - 1/2]\, du = 1 + o(1))f_0(k_\alpha)k_\alpha/j, \tag{4.6}$$

by partial integration. On the other hand, an integration by parts shows that

$\cdots, u$) and variance σ^2 (we use a convenient notation for the example, rather than that introduced in Section 1). The problem is to test $H_0: \mu_1 = \mu_2 = \cdots = \mu_u = 0$, and a design d in Δ is a specification of nonnegative integers n_i whose sum is N. For any such d, we denote by $M(d)$ the set of i for which $n_i > 0$; by $k(d)$, the number of integers in $M(d)$; by τd, the design associated with the values $n_i = n^*_{\tau(i)}$ when d is associated with the values $n_i = n^*_i$, where τ is any element of the symmetric group S_u on u symbols; by δ_d, the design in Δ_R which assigns probability $1/u!$ to each τd for τ in S_u; by $f_{d,\alpha}$ the test associated with δ_d which is obtained by using the appropriate F-test of size α with whatever τd is chosen by δ_d. We shall also use the symbol $a_\phi(c)$ of (2.2), with $\psi(\mu) = \sum_1^u \mu_i^2$, and shall denote by a'_ϕ its derivative with respect to c at $c = 0$. We shall also use the symbols $g_{ij}(\alpha)$ introduced in (1.3). Our result, which implies that the "symmetrical" design associated with $k(d) = u$ and all n_i equal (or as nearly so as possible) is not L_α-optimum in Δ_R, and that the δ_d associated with the d for which $n_1 = N$ (this δ_d chooses each i with probability $1/u$ and takes all Y_{ij} with the chosen i) is locally best among the δ_d, is the following:

THEOREM 4.1. *For every d, α, and c,*

$$a_{F_{d,\alpha}}(c) \leqq a_{f_{d,\alpha}}(c); \tag{4.1}$$

$a'_{f_{d,\alpha}}$ is strictly decreasing in $k(d)$, and the same is true of $a_{f_{d,\alpha}}(c)$ for all c in some neighborhood of $c = 0$.

PROOF. (4.1) is trivial, and we proceed to the rest of the proof. The numerator $t'_d V_d^{-1} t_d$ of $F_{d,\alpha}$ is of course

$$U_d = \sum_{i \varepsilon M(d)} n_i \left(\sum_{j=1}^{n_i} Y_{ij}/n_i \right)^2,$$

and U_d/σ^2 has a χ^2 distribution with $k(d)$ d.f. and non-central parameter

$$\sum_{i \varepsilon M(d)} n_i \mu_i^2/\sigma^2.$$

The denominator of $F_{d,\alpha}$ has $N - k(d)$ d.f. Write $\lambda = \sum_1^u \mu_i^2/\sigma^2$. From (1.3) we have, as $\lambda \to 0$,

$$\begin{aligned} \beta_{f_{d,\alpha}}(\mu, \sigma^2) &= \sum_{\tau \varepsilon S_u} \beta_{F_{\tau d,\alpha}}(\mu, \sigma^2)/u! \\ &= \sum_{\tau \varepsilon S_u} [\alpha + g_{k(d),N-k(d)}(\alpha) \sum_{\tau \varepsilon S_u} n_{\tau(i)}\mu_i^2/\sigma^2 + 0(\lambda^2)]/u! \\ &= \alpha + g_{k(d),N-k(d)}(\alpha) \sum_i \left(\sum_\tau n_{\tau(i)}/u!\right)\mu_i^2/\sigma^2 + 0(\lambda^2) \\ &= \alpha + \frac{N}{u} g_{k(d),N-k(d)}(\alpha)\lambda + 0(\lambda^2). \end{aligned} \tag{4.2}$$

The desired conclusion now follows from Lemma 4.1.

Existing tables and charts of the power functions of the F-test and χ^2-test are presented in such forms (in terms of $\sqrt{\lambda/(k(d) + 1)}$, usually in inverted form and with wide spacing of arguments) as to make accurate comparisons of the

$\beta_{f_{d,\alpha}}$ difficult. This difficulty is made the worse by the fact that $\beta_{f_{d,\alpha}}$ is not (with an obvious exception) constant on the contour λ = constant, making it somewhat of a task to obtain $a_{f_{d,\alpha}}(c)$. It is not true, as might be supposed, that this minimum power on the contour λ = constant is always attained for a μ with all components equal, or else is always attained for a μ with all components except one equal to zero. To see this, consider the problem of Section 1 ($N = u = 2$, σ^2 known). Let C_α be the value such that, if Y is a normal random variable with 0 mean and unit variance, then $P\{ \mid Y \mid > C_\alpha\} = \alpha$. A direct computation of the power function of δ near $\lambda \equiv \mu_1^2 + \mu_2^2 = 0$ yields

$$\beta_\delta(\mu) = \alpha + \frac{C_\alpha \exp(-C_\alpha^2/2)}{2\sqrt{2\pi}} \cdot \{2(\mu_1^2 + \mu_2^2) + (C_\alpha^2 - 3)(\mu_1^4 + \mu_2^4)/3 + O(\lambda^3)\}. \tag{4.3}$$

Hence, when c is sufficiently small, the minimum of $\beta_\delta(\mu)$ on the contour $\lambda = c$, neglecting the term $O(\lambda^3)$, is located at $\mu_1 = \sqrt{c}$, $\mu_2 = 0$ (or $\mu_2 = \sqrt{c}$, $\mu_1 = 0$) if $C_\alpha \leqq \sqrt{3}$ and at $\mu_1 = \mu_2 = \sqrt{c/2}$ if $C_\alpha \geqq \sqrt{3}$. When we include terms of higher order in μ, it is no longer even evident that the minimum must be attained at one of these two values of μ.

We see from (4.3) that $g_{1,\infty}(\alpha) = (2\pi)^{-\frac{1}{2}} C_\alpha \exp(-C_\alpha^2/2)$ and it is not hard to show that $g_{2,\infty}(\alpha) = -\alpha(\log \alpha)/2$ (see [12], equation (6.27), where λ is our $\lambda/2$). Thus, a comparison of $a'_{f_{d,\alpha}}$ for $k(d) = 1$ and 2 is given in this example by the following table:

α	$g_{1,\infty}(\alpha)$	$g_{2,\infty}(\alpha)$
01	.037	.023
.05	.114	.075
.10	.175	.115
.20	.225	.161
.30	.242	.181
.50	.214	.173
.90	.050	.047

The following lemma shows that, as $\alpha \to 0$, the ratio of the second to third column above goes to 2 and, more generally, that $g_{i,\infty}(\alpha)/g_{j,\infty}(\alpha) \to j/i$ (this gives a comparison of the various δ_d for general N and u and for various $k(d)$ when σ^2 is known, as $\alpha \to 0$; see Lemma 4 3 for the case when σ^2 is unknown):

LEMMA 4.2. *As* $\alpha \to 0$,

$$g_{j,\infty}(\alpha) = -[1 + o(1)]\alpha(\log \alpha)/j. \tag{4.5}$$

PROOF. Fix j. Let k_α be such that if Y is a random variable with central χ^2 distribution with j d.f., then $P\{Y > k_\alpha\} = \alpha$. Let f_λ be the χ^2 density function with j d.f. and non-central parameter λ. A simple calculation shows that $df_\lambda(u)/d\lambda$ at $\lambda = 0$ is $f_0(u)[(u/2j) - 1/2]$. Hence, as $k_\alpha \to \infty$,

$$g_{j,\infty}(\alpha) = \int_{k_\alpha}^{\infty} f_0(u)[(u/2j) - 1/2]\, du = 1 + o(1))f_0(k_\alpha)k_\alpha/j, \tag{4.6}$$

by partial integration. On the other hand, an integration by parts shows that

$= 2f_0(k_\alpha)[1 + o(1)]$ as $k_\alpha \to \infty$, and hence that $k_\alpha = -2[1 + o(1)] \log \alpha$. is completes the proof.

IB. CASE II. We again treat the case where σ^2 is unknown, the other case ng handled similarly (mainly, use Lemma 4.2 for Lemma 4.3). We first prove o simple lemmas.

LEMMA 4.3. *As* $\alpha \to 0$,

$$g_{ji}(\alpha) = i\alpha/2j + o(\alpha). \tag{7}$$

(This does not contradict (4.5), since j is fixed in (4.7).)

PROOF: Fix j and i. Let h_α be such that if Y has a central F-distribution with nd i d.f., then $P\{Y > h_\alpha\} = \alpha$. Let G_λ be the F density function with j and .f. and non-central parameter λ. From [12], equation (6.29) (with λ there laced by our $\lambda/2$), it is easy to compute that $dG_\lambda(u)/d\lambda$ at $\lambda = 0$ is $G_0(u)$ $+ i) u/j(1 + u) - 1]/2$. Hence, as $k_\alpha \to \infty$,

$$g_{ji}(\alpha) = \frac{1}{2}\int_{k_\alpha}^{\infty} G_0(u)\left[\frac{i}{j} - \frac{j+i}{j}\cdot\frac{1}{1+u}\right]du = i\alpha/2j + o(\alpha). \tag{8}$$

In the next lemma, we use the following notation: n_i $(i = 1, \cdots, u)$ are again nnegative integers with sum N. S_u is the symmetric group on u symbols and, τ in S_u, $\bar{\mu}(\tau) = N^{-1}\sum_i n_{\tau(i)}\mu_i$; finally, $\bar{\mu} = u^{-1}\sum_i \mu_i$.

LEMMA 4.4. *For all* $u > 1$, μ, *and* N,

$$\sum_{\tau \in S_u}\sum_i n_{\tau(i)}(\mu_i - \bar{\mu}(\tau))^2 = u(u-2)![N - N^{-1}\sum_i n_i^2]\sum_i(\mu_i - \bar{\mu})^2. \tag{9}$$

PROOF. Since

$$\sum_{\tau \in S_u} n_{\tau(i)}^2 = (u-1)!\sum_i n_i^2 \tag{10}$$

d, for $i \neq j$,

$$\sum_{\tau \in S_u} n_{\tau(i)}n_{\tau(j)} = (u-2)!\sum_{i\neq j} n_i n_j = (u-2)![N^2 - \sum_i n_i^2], \tag{11}$$

have

$$\begin{aligned} N^2\sum_{\tau\in S_u}\bar{\mu}(\tau)^2 &= \sum_{i,j}\mu_i\mu_j\sum_{\tau\in S_u} n_{\tau(i)}n_{\tau(j)} \\ &= (u-1)!\sum_i n_i^2\sum_j \mu_j^2 \\ &\quad + (u-2)![N^2 - \sum_i n_i^2][u^2\bar{\mu}^2 - \sum_j \mu_j^2]. \end{aligned} \tag{12}$$

so,

$$\sum_{\tau\in S_u}\sum_i n_{\tau(i)}\mu_i^2 = \sum_i \mu_i^2\sum_\tau n_{\tau(i)} = (u-1)!N\sum_i \mu_i^2. \tag{13}$$

quations (4.12) and (4.13), together with

$$\sum_\tau\sum_i n_{\tau(i)}(\mu_i - \bar{\mu}(\tau))^2 = \sum_\tau\sum_i n_{\tau(i)}\mu_i^2 - N\sum_\tau \bar{\mu}(\tau)^2, \tag{14}$$

ve (4.9).

The maximum for fixed $k(d)$ of the factor in square brackets on the right side of (4.9) will of course be nondecreasing in $k(d)$. It is the factor $g_{k(d)-1,h_d(\alpha)}$ which will increase rapidly enough as $k(d)$ is decreased to more than make up for the decrease in this term in brackets.

We are now ready to give our nonoptimality result in several illustrative examples of Case II, including those of Section 3C(1) and 3C(2). In all of these examples we ignore the divisibility properties; considerations when the design does not "divide up" properly (e.g., when $k(d)$ does not divide N in Example (1) below) are messier and their consideration does not help in the understanding of the phenomenon we are illustrating; thus, we shall assume whatever divisibility properties of N are needed to make our examples simple.

(1). *One-way analysis of variance.* In our first and simplest example, the setup is that of Section 4A, except that we now are testing $\mu_1 = \cdots = \mu_u$, and the appropriate F-tests are changed accordingly. Our result has the same implication as that stated just above Theorem 4.1, except that it now holds only when α is sufficiently small, and the optimum δ chooses each pair (i, j) $(i \neq j)$ with equal probability and sets $n_i = n_j = N/2$.

THEOREM 4.21. *For every d, α, and c, (4.1) holds; for fixed $k(d)$, $a'_{f_{d,\alpha}}$ is strictly decreasing in $\sum_i n_i^2$, attaining its maximum for $n_1 = \cdots = n_{k(d)} = N/k(d)$, for this choice of the n_i and for all α in some neighborhood of 0, $a'_{f_{d,\alpha}}$ is strictly decreasing in $k(d)$ for $k(d) > 1$; the results just stated for $a'_{f_{d,\alpha}}$ hold also for $a_{f_{d,\alpha}}(c)$ for all c in some neighborhood of* 0.

PROOF. From Lemma 4.4 and an argument like that of (4.2), we have, setting $\lambda = \sum_i(\mu_i - \bar{\mu})^2/\sigma^2$,

$$\beta_{f_{d,\alpha}}(\mu, \sigma^2) = \alpha + g_{k(d)-1,N-k(d)}(\alpha)(u-1)^{-1}(N - N^{-1}\sum_i n_i^2)\lambda + O(\lambda^2). \tag{4.15}$$

When $n_1 = \cdots = n_{k(d)} = N/k(d)$, the ratio of values of $a'_{f_{d,\alpha}}$ corresponding to two values k and k' of $k(d)$ with $1 < k < k'$ is thus

$$\frac{g_{k-1,N-k}(\alpha)(1 - 1/k)}{g_{k'-1,N-k'}(\alpha)(1 - 1/k')}; \tag{4.16}$$

as $\alpha \to 0$, by Lemma 4.3, this ratio approaches

$$\frac{(N-k)/k}{(N-k')/k'} > 1, \tag{4.17}$$

completing the proof.

For a numerical example, suppose $N = 6$, $u = 3$, with σ^2 known. Comparing the δ_d's for which $k = 2$ and $k' = 3$, we see that $(1 - 1/k)/(1 - 1/k') = \frac{3}{4}$; thus, the ratio of the two $a'_{f_{d,\alpha}}$ in this example is $\frac{3}{4}$ times the ratio of second to third column in the table above Lemma 4.2. For $\alpha < .3$, then, the design with $k(d) = 2$ is locally better than that with $k(d) = 3$, in this example.

(2). *Several-way analysis of variance.* With or without interactions, the considerations are very similar to those of Example (1), and we omit them.

(3). *One-way heterogeneity.* In the setting described in Section 3A, suppose for

b, k, and u that BBD's exist for two possible choices u_1 and u_2 of the "num- of treatments" to be tested, say for u_1 and u_2 with $1 < u_1 < u_2 \leqq u$. Let $= 1, 2)$ be the design which uses the BBD with parameters b, k, and u_i to the hypothesis $\mu_1 = \cdots = \mu_{u_i}$, and let δ_{d_i} be the corresponding randomized gn which replaces the subscripts $1, \cdots, u_i$ here by $\tau(1), \cdots, \tau(u_i)$ with ability $1/u!$ for each τ (or, which is the same thing, which chooses each of possible subsets of u_i treatments with equal probability). Otherwise, we use same notation as in Example (1) of this section.

or any design setting, the parameter of the non-central χ^2 variable $t_d' V_d^{-1} t_d/\sigma^2$ $_dR\mu)' V_d^{-1}(Q_dR\mu)$, and by Lemma 2.3 and equation (3.1) this reduces in the of a BBD d^* with parameters b, k, and u to

$$3) \qquad [r_{d^*1} - (\lambda_{d^*11} - \lambda_{d^*12})/k] \sum_i (\mu_i - \bar{\mu})^2/\sigma^2.$$

the sake of arithmetical simplicity only, suppose that k/u_i is either an integer < 1 (the phenomenon to be studied persists without this assumption). n, for $d^* = d_i$, the term in square brackets in (4.18) is easily computed to be

$$9) \qquad f(u_i) = \begin{cases} b(k-1)/(u_i - 1) & \text{if } k/u_i \leqq 1, \\ bk/u_i & \text{if } k/u_i \geqq 1. \end{cases}$$

g now the counterpart of (4.18) for the designs d_i and the fact that, for $\cdots = n_{u_q} = 1$ and all other $n_j = 0$, (4.9) becomes

$$)) \qquad \sum_{\tau \varepsilon S_u} \sum_i n_{\tau(i)}(\mu_i - \bar{\mu}(\tau))^2/u! = (u-1)^{-1}(u_q - 1) \sum_{i=1} (\mu_i - \bar{\mu})^2,$$

btain, corresponding to (4.16),

$$1) \qquad \frac{a'_{f_{d_1,\alpha}}}{a'_{f_{d_2,\alpha}}} = \frac{g_{u_1-1, bk-u_1-b+1}(\alpha)(u_1 - 1)f(u_1)}{g_{u_2-1, bk-u_2-b+1}(\alpha)(u_2 - 1)f(u_2)}.$$

Lemma 4.3, as $\alpha \to 0$ this ratio approaches

$$2) \qquad \frac{(bk - u_1 - b + 1)f(u_1)}{(bk - u_2 - b + 1)f(u_2)}.$$

trivial to verify that $(bk - u - b + 1)f(u)$ is strictly decreasing in u for 1, so that the expression of (4.22) is >1. Thus, we have proved

HEOREM 4.22. *For fixed* b, k, u *and all* α *in some neighborhood of* 0, $a'_{f_{d_i,\alpha}}$ *is tly decreasing in* u_i *for* $i > 1$; *the same is true for* $a_{f_{d_i,\alpha}}(c)$ *for all* c *in some hborhood of* 0.

his result implies that, if k is even, the locally best δ_{d_i} is that which chooses a pair of treatments with equal probability and assigns each of the two chosen tments to $k/2$ of the plots in *every* block.

4). *Two-way heterogeneity*. Using (3.2) in place of (3.1), the analogue of orem 4.22 can be proved for the YS design by an argument very similar to t of Example (3) just above, and which we therefore omit. One can even give

an example of the lack of optimality of the YS in Δ_R without resorting to this analysis: for the case $k_1 = 2$, $k_2 = 3$, $u = 3$, the usual YS gives no d.f. to error, while the design which chooses two treatments at random and assigns each treatment to three plots, at least once in each row and column (full symmetry is impossible here) is uniformly more powerful for all α and all alternatives.

(5) *Other examples.* Examples like those mentioned in Section 3C (3) can be considered similarly, with analogous results. In particular, a trivial example in the case of a higher LS has already been mentioned in the first paragraph of Section 1.

5. Remarks and extensions. We list a few of the variants of the examples considered in this paper for which similar results hold, and make a few comments on questions which arise in connection with the paper, some of which present unanswered research problems.

1. A few of the other problems to which modifications of our method apply have been mentioned in Section 3C, and some of these will be considered elsewhere. Some such results hold under various non-normal probability laws (the point of the results of Section 4 is not merely that they hold for *many* models, but that they hold for the simplest, classical, normal model). Of course, a design which is optimum for one model may fail to be optimum for another, and vice versa; in particular, the results are obviously sensitive to change in the function ψ (even to changes to other quadratic forms and for a fixed d, as indicated in Section 2). Optimality criteria can be altered in other ways; e.g., one can consider $M_{\alpha,c,\sigma}$-optimality, in imitation of 2A(c). The extent of completeness of nonoptimality results like those on the higher LS design (first paragraph of Section 1) and YS design (Section 4B(4)) obviously depends on whether or not σ^2 is known. The results for Model II and certain mixed models of the analysis of variance differ considerably from those for the model considered herein, since the dependence of the power function on the design (and on the test, for a fixed design) is so different; however, similar methods can be used there.

2. Besides changing the model, one can also change the decision space. From the examples cited just above regarding higher LS and YS designs, it is clear that *nonoptimality* results for some classical symmetrical designs hold for many decision problems. For normal and certain nonparametric point estimation problems, the discussion of [2] and [3] indicates why Section 3 yields *optimality* results (these actually hold for many weight functions other than squared error). Another typical estimation result is contained in the fact that the designs d^* of Theorems 3.1 and 3.2 maximize the trace of V_d^{-1} and that V_{d^*} is a multiple of the identity; from these it follows at once that *average variance of* t_d ($=$ trace of $\sigma^2 V_d/(u-1)$) *is a minimum for* d^*. However, the results of Section 4 are meaningless for many common weight functions, since V_δ is not the covariance matrix of b.l.e.'s. Similar results hold for some interval estimation problems; for estimating $\psi(\mu)/\sigma^2$ (e.g., in "multiple comparison" problems), Section 4 is now sometimes relevant. Multiple classification and ranking problems can be treated in like

anner. Of course, a D-optimum design minimizes the approxiate *generalized riance* in point estimation problems.

3. As we have mentioned, nonoptimality results like those of Section 4 do not epend on the nonrandomized design being symmetrical. Much more difficult is e problem of characterizing optimum designs in the sense of Section 3 when ere is no appropriate symmetry. (Even the considerations of Sections 3B(2) id 4B(3 and 4) become messier without the restrictions on k_i/u and k/u; it ould be nice if neat proofs could be given in such cases.) It seems often to be ue that a design which is "closest to being symmetrical" in an appropriate nse (e.g., note the dependence on $\sum n_i^2$ in Theorem 4.21) is optimum, but the gebra involved in proving this can be tedious. Problems like that cited in the ext to last paragraph of Section 3C(3) can be similarly unwieldy under heter-scedasticity. In connection with a general symmetry-invariance approach like at mentioned below (1.3), we note that appropriate symmetry of X_d is useful as partial sufficient condition for some optimality results, but that appropriate mmetry of $X_d'X_d$ is what is really relevant (for the functions ψ we have con-dered).

4. We have mentioned in Section 2 some of the difficulties present in verifying - (or sometimes L-) optimality. If b_d is not a constant for d in Δ', or if ran-omized designs are considered, this difficulty is increased by the nonconstancy the d.f. for $\bar{S}_d$, etc. (We have not considered here a thorough investigation of e optimality properties of the procedures δ_d of Section 4). The difficulty en-untered in connection with M-optimality in the nonconstancy of the power inctions of competing tests on appropriate contours also manifests itself when e tries to find a *most stringent* design (the "envelope power function" being otained by taking the supremum of β_ϕ over all ϕ in $H_d(\alpha)$ and all d in Δ or $_R$). The method of invariance used to prove 2A(f) cannot even supply a start ere, and the method of [6] or [7] used to prove 2A(c) yields no analogue here here d is not fixed. Thus, even in such a simple example as that of Section 2B, e stringency problem seems extremely difficult.

It is interesting to note that the δ_d of Section 4 lack a "consistency" property if $(d) < r$, in that $a_{f_{d,\alpha}}(c)$ does not approach 1 as $c \to \infty$ (in fact, it is easy to e that the μ for which one component of $R\mu$ is $\sigma\sqrt{c}$ and all others are 0 is symptotically worst on the contour $\psi(u)/\sigma^2 = c$ as $c \to \infty$, giving power ap-roaching $[k(d) + (r - k(d))\alpha]/r$). Nevertheless, the question remains open as to hether any of these δ_d, or some other design and associated test which lacks this onsistency property, is nevertheless most stringent.

The reader will not find it difficult in considerations like those of Section 3B to ipply the details which show, in some problems, that the D-optimum (or L- r E-optimum) design is unique. When uniqueness is not present (e.g., for some and ϵ, both designs in Section 2B will be L-optimum), questions of global dmissibility arise. A related problem is to look not at a fixed contour or family of ontours in the manner of Section 2, but rather to characterize complete classes designs in the manner of [3]; in such considerations, especially for problems of

testing hypotheses, Section 4 shows that results like those of [3] must be altered if Δ_R is considered rather than Δ.

Finally, we may remark that, for a fixed d, the problem of characterizing an L_α-optimum test is unsolved; the generalized Neyman-Pearson Lemma does not seem to yield explicit results easily, although it is not difficult to show that an L_α-optimum test is obtained by replacing the numerator of the F-test by some other quadratic form.

REFERENCES

[1] A. WALD, "On the efficient design of statistical investigations," *Ann Math. Stat*, Vol. 14 (1943), pp. 134–140.

[2] S. EHRENFELD, "On the efficiency of experimental designs," *Ann. Math. Stat.*, Vol. 26 (1953), pp. 247–255.

[3] S EHRENFELD, "Complete class theorems in experimental design," *Third Berkeley Symposium on Probability and Statistics*, Vol 1, University of California Press, 1946

[4] R. C BOSE, "Least Squares Aspects of Analysis of Variance" (mimeographed notes), Institute of Statistics, North Carolina

[5] E L. LEHMANN, "Notes on Testing Hypotheses" (mimeographed), Associated Student Store, University of California, Berkeley

[6] A. WALD, "On the power function of the analysis of variance test," *Ann. Math Stat*, Vol. 13 (1942), pp 434–439, and Vol 15 (1944), pp 330–333

[7] J. WOLFOWITZ, "The power of the classical tests associated with the normal distribution," *Ann Math. Stat*, Vol. 20 (1949), pp 540–551.

[8] P. L. HSU, "Analysis of variance from the power function standpoint," *Biometrika*, Vol 32 (1941), p. 62.

[9] G A HUNT AND C. STEIN, "Invariant tests," unpublished.

[10] J. KIEFER, "Invariance, sequential minimax estimation, and continuous time processes," *Ann. Math. Stat.*, Vol. 28 (1957).

[11] S. ISAACSON, "On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters," *Ann. Math. Stat*, Vol. 22 (1951), pp. 217–234.

[12] H. B. MANN, *Analysis and Design of Experiments*, Dover, New York, 1949.

DISTINGUISHABILITY OF SETS OF DISTRIBUTIONS

(The case of independent and identically distributed chance variables)

By Wassily Hoeffding[1] and J. Wolfowitz[2]

University of North Carolina and Cornell University

1. Introduction. Suppose it is desired to make one of two decisions, d_1 and d_2, on the basis of independent observations on a chance variable whose distribution F is known to belong to a set $\mathfrak{F}$. There are given two subsets $\mathfrak{G}$ and $\mathfrak{H}$ of $\mathfrak{F}$ such that decision $d_1(d_2)$ is strongly preferred if F is in $\mathfrak{G}$ ($\mathfrak{H}$). Then it is reasonable to look for a test (decision rule) which makes the probability of an erroneous decision small when F belongs to $\mathfrak{G}$ or $\mathfrak{H}$, and at the same time exercises some control over the number of observations required to reach a decision when F is in $\mathfrak{F}$ (not only in $\mathfrak{G}$ or $\mathfrak{H}$).

This paper is concerned with criteria that enable us to decide whether, for given sets $\mathfrak{F}$, $\mathfrak{G}$, and $\mathfrak{H}$, there exists a test of the described type. More precisely, we shall consider several classes of tests, such as the class of all fixed sample size tests, or the class of all tests which terminate with probability one whenever F is in $\mathfrak{F}$. Thus the restriction to tests in one of these classes is equivalent to imposing some sort of control, of a purely qualitative nature, on the sample size. We then shall try to find necessary and/or sufficient conditions for the existence of a test in a given class which makes the maximum error probability in $\mathfrak{G} \cup \mathfrak{H}$ less than any preassigned positive number.

If such a test exists, we shall say that the sets $\mathfrak{G}$ and $\mathfrak{H}$ are distinguishable[3] in the given class $\mathfrak{I}$ of tests. If $\mathfrak{I}$ is the class of all fixed sample size tests, the distinguishability of $\mathfrak{G}$ and $\mathfrak{H}$ in $\mathfrak{I}$ is equivalent to the existence of what has been called a uniformly consistent sequence of tests for testing $F \varepsilon \mathfrak{G}$ against $F \varepsilon \mathfrak{H}$.

The sets $\mathfrak{G}$ and $\mathfrak{H}$ will be called indistinguishable in $\mathfrak{I}$ if for any test in $\mathfrak{I}$ the sum of the maximum error probability in $\mathfrak{G}$ and the maximum error probability in $\mathfrak{H}$ is at least one. (There always exists a trivial test for which this sum is equal to one.) In section 2 it will be shown that, with the present restriction to sequences of independent and identically distributed chance variables, two sets are either distinguishable or indistinguishable in any of the classes $\mathfrak{I}$ which we shall consider.

Received May 20, 1957.

[1] The research of this author was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract No. AF 18(600)-458. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] The research of this author was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract No. AF 18(600)-685. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[3] In [3] the term distinguishable was used in another sense.

Since we confine ourselves to tests based on a sequence $X_1, X_2, \cdots$ of independent, identically distributed chance variables, we may restrict ourselves to sequential tests. A sequential test is determined by the sample size function N and the terminal decision function ϕ, and will be denoted by (N, ϕ). Here N is a chance variable whose values are the non-negative integers and $+\infty$, and whose conditional distribution, given any sequence $\mathbf{x} = (x_1, x_2, \cdots)$ of possible values of $X_1, X_2, \cdots$, is such that the probability of $N \leqq n$ does not depend on $x_{n+1}, x_{n+2}, \cdots$, for all non-negative integers n. The function ϕ is a function of $(x_1, \cdots, x_N)$ whose values range from 0 to 1. The test (N, ϕ) consists in taking one observation on each of the first N chance variables of the sequence, finding the corresponding value of ϕ, and making decision d_1 or d_2 with respective probabilities $1 - \phi$ and ϕ. The function ϕ and the conditional distribution function of N given $\mathbf{x}$ are always understood to be measurable on the appropriate σ-field. The sample size function N and the terminal decision function ϕ are said to be non-randomized if the respective functions $P[N \leqq n \mid \mathbf{x}]$ and $\phi(\mathbf{x})$ take on the values 0 and 1 only. A test (N, ϕ) will be called non-randomized if both N and ϕ are non-randomized.

We use the term distribution synonymously with probability measure. The set $\mathfrak{F}$ consists of distributions on a fixed σ-field $\mathfrak{A}$ of subsets of a space $\mathfrak{X}$. Unless we state otherwise, we shall assume that $\mathfrak{X}$ is a k-dimensional Euclidean space and $\mathfrak{A}$ the k-dimensional Borel field. A distribution on $\mathfrak{A}$ will then be called a k-dimensional or k-variate distribution If F is a distribution on $\mathfrak{A}$, we denote by $F[A]$ the probability of the set $A \varepsilon \mathfrak{A}$ and by $F(x) = F(x^{(1)}, \cdots, x^{(k)})$, $x \varepsilon \mathfrak{X}$, the associated distribution function, that is, $F(x) = F[\{y \mid y^{(1)} \leqq x^{(1)}, \cdots, y^{(k)} \leqq x^{(k)}\}]$. With the usual definition (see [5])[4] of the distribution of a sequence $\mathbf{X} = (X_1, X_2, \cdots)$ of independent chance variables with identical marginal distribution F, we denote by $P_F[B]$ the probability of a measurable set B in the range of $\mathbf{X}$, and write $E_F\psi$ for the expected value of a function ψ of $\mathbf{X}$.

According to our definitions, the probability of an erroneous decision when test (N, ϕ) is used is equal to $E_F\phi$ if $F \varepsilon \mathcal{G}$, and to $E_F(1 - \phi)$ if $F \varepsilon \mathcal{H}$ Thus the sets $\mathcal{G}$ and $\mathcal{H}$ are distinguishable in a class $\mathfrak{I}$ of tests if and only if for every $\epsilon > 0$ there exists a test (N, ϕ) in $\mathfrak{I}$ such that $E_F\phi < \epsilon$ for $F \varepsilon \mathcal{G}$ and $E_F(1 - \phi) < \epsilon$ for all $F \varepsilon \mathcal{H}$.

2. Modes of distinguishability. We shall be concerned with the distinguishability of two sets of distributions in various classes $\mathfrak{I}$ of tests, which are defined in terms of properties of the distribution of the sample size function N. Some classes of particular interest are the following.

$\mathfrak{I}_0$: $P_F[N < \infty] = 1$ if $F \varepsilon \mathfrak{F}$

$\mathfrak{I}_1(r)$: $E_F N^r < \infty$ if $F \varepsilon \mathfrak{F}$ $(r > 0)$.

$\mathfrak{I}_1$: $E_F N^r < \infty$ for all $r > 0$ if $F \varepsilon \mathfrak{F}$.

$\mathfrak{I}_2$: $E_F e^{tN} < \infty$ for some $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.

$\mathfrak{I}_3$: $\max(N) < \infty$.

---

[4] The numbers in square brackets refer to the bibliography listed at the end.

It will be noted that each of the successive classes contains the one following. Some classes of obvious interest have been omitted because, for the purposes of our investigation, they are equivalent to some of the classes listed above. Thus if two sets are distinguishable in one of the classes $\mathfrak{J}_0, \cdots, \mathfrak{J}_3$, they are also distinguishable in the corresponding subclass which contains only the non-randomized tests; this follows from Theorem 2.1 below. If two sets are distinguishable in $\mathfrak{J}_3$ (the class of "truncated" sequential tests), they are clearly distinguishable in the class of all fixed sample size tests; for if (N, ϕ) is any test in $\mathfrak{J}_3$, and we put $N' = \max (N)$, $\phi' = E[\phi \mid x]$, then (N', ϕ') is a fixed sample size test such that $E_F\phi' = E_F\phi$ for all F.

In view of the importance of the two extreme classes, $\mathfrak{J}_0$ and $\mathfrak{J}_3$, we shall use the following terms. If two sets of distributions are distinguishable (indistinguishable) in $\mathfrak{J}_0$, they will be called distinguishable $(\mathfrak{F})$(indistinguishable $(\mathfrak{F})$). If two sets are distinguishable (indistinguishable) in $\mathfrak{J}_3$, we shall say that they are finitely distinguishable (finitely indistinguishable).

The classes $\mathfrak{J}_i$ have been defined in terms of the set $\mathfrak{F}$ to which the distribution of X_j is assumed to belong (without displaying $\mathfrak{F}$ in the notation). It may be of interest to consider also the corresponding classes where $\mathfrak{F}$ is replaced by some subset of $\mathfrak{F}$ (compare Lemma 4.1 in section 4). It will be clear that Theorems 2.1 and 4.1 below can be immediately extended to such classes.

Our list does not contain the subclass of $\mathfrak{J}_1(r)$ where $E_F N^r$ is bounded for $F \varepsilon \mathfrak{F}$, nor the subclass of $\mathfrak{J}_0$ where $P_F[N > n] \to 0$ as $n \to \infty$, uniformly for $F \varepsilon \mathfrak{F}$. The reason for this omission is that two sets $\mathcal{G}$ and $\mathcal{H}$ which are distinguishable in one of these classes are finitely distinguishable. This follows from the following fact: If (N, ϕ) is a test such that $P_F[N > n] \to 0$ as $n \to \infty$, uniformly for $F \varepsilon \mathcal{G} \cup \mathcal{H}$, then for every $\epsilon > 0$ there exists a test (N', ϕ') such that $\max (N') < \infty$ and $| E_F\phi' - E_F\phi | < \epsilon$ for all $F \varepsilon \mathcal{G} \cup \mathcal{H}$. This is so since, by our assumption, we can choose an integer $n = n(\epsilon)$ such that $P_F[N > n] < 2 \epsilon$ for all $F \varepsilon \mathcal{G} \cup \mathcal{H}$, and the test (N', ϕ') defined by

$$\phi' = \phi, \qquad N' = N \text{ if } N \leqq n; \qquad \phi' = \tfrac{1}{2}, \qquad N' = n \text{ if } N > n$$

has the stated property.

Let $\mathfrak{J}$ be any class of tests. If $\Phi = \Phi(\mathfrak{J})$ denotes the class of all terminal decision functions ϕ of the tests in $\mathfrak{J}$, the statement that $\mathcal{G}$ and $\mathcal{H}$ are distinguishable in $\mathfrak{J}$ can be expressed by the equation

$$\sup_{\phi \varepsilon \Phi} \inf_{G \varepsilon \mathcal{G}, H \varepsilon \mathcal{H}} (E_H\,\phi - E_G\,\phi) = 1. \tag{2.1}$$

Whenever $\mathfrak{J}$ contains a trivial test such that $\phi = \text{const}$, the left side of (2.1) is at least zero. Let us say that a test in $\mathfrak{J}$ is nontrivial for distinguishing between $\mathcal{G}$ and $\mathcal{H}$ if $\sup_{G \varepsilon \mathcal{G}} E_G\phi < \inf_{H \varepsilon \mathcal{H}} E_H\phi$. Thus the left side of (2.1) is positive if and only if $\mathfrak{J}$ contains a nontrivial test for distinguishing between $\mathcal{G}$ and $\mathcal{H}$. The following theorem shows that if $\mathfrak{J}$ is one of the classes $\mathfrak{J}_0, \cdots, \mathfrak{J}_3$ (or one of the "equivalent" classes mentioned above), then the existence in $\mathfrak{J}$ of a nontrivial

test for distinguishing between $\mathcal{G}$ and $\mathcal{H}$ is sufficient for $\mathcal{G}$ and $\mathcal{H}$ to be distinguishable in $\mathfrak{J}$, and even in the class $\mathfrak{J}'$ which consists of the non-randomized tests in $\mathfrak{J}$. The special case of the theorem where $\mathfrak{J}$ is the class of all non-randomized fixed sample size tests is contained in a lemma of Berger [1] (which is there attributed to Bernoulli).

We denote by Φ and Φ' the classes of the terminal decision functions of the tests in $\mathfrak{J}$ and $\mathfrak{J}'$, respectively.

THEOREM 2.1. *If $\mathfrak{J}$ is one of the classes $\mathfrak{J}_0, \cdots, \mathfrak{J}_3$, then*

$$\sup_{\phi \varepsilon \Phi} \inf_{G \varepsilon \mathcal{G}, H \varepsilon \mathcal{H}} (E_H \phi - E_G \phi) > 0 \tag{2.2}$$

implies

$$\sup_{\phi \varepsilon \Phi'} \inf_{G \varepsilon \mathcal{G}, H \varepsilon \mathcal{H}} (E_H \phi - E_G \phi) = 1 \tag{2.3}$$

Hence

$$\sup_{\phi \varepsilon \Phi} \inf_{G \varepsilon \mathcal{G}, H \varepsilon \mathcal{H}} (E_H \phi - E_G \phi) = 0 \text{ or } 1. \tag{2.4}$$

For the proof of Theorem 2.1 we require the following

LEMMA. *If $\mathfrak{J}$ is one of the classes $\mathfrak{J}_0, \cdots, \mathfrak{J}_3$, and (N, ϕ) is in $\mathfrak{J}$, then for every $\epsilon > 0$ there is a test (N', ϕ') in $\mathfrak{J}$ such that N' is non-randomized and $|\phi' - \phi| < \epsilon$*

PROOF. Let N' be the least integer $n \geqq 1$ such that

$$P[N > n \mid \mathbf{x}] < \epsilon.$$

Define ϕ' by

$$\phi' = E[\phi \mid N \leqq n, \mathbf{x}] \text{ if } N' = n, \qquad n = 1, 2, \cdots.$$

Thus (N', ϕ') is a test, and N' is non-randomized.

We have for every $n \geqq 1$

$$P[N' > n] = P\{P[N > n \mid \mathbf{x}] \geqq \epsilon\} \leqq \epsilon^{-1} EP[N > n \mid \mathbf{x}]$$
$$= \epsilon^{-1} P[N > n].$$

Since for any increasing function h on the nonnegative integers

$$Eh(N) = h(0) + \sum_{n=0}^{\infty} [h(n+1) - h(n)] P[N > n],$$

it follows that if N satisfies the condition for any of the classes $\mathfrak{J}_0, \cdots, \mathfrak{J}_3$, so does N'. Hence (N', ϕ') is in $\mathfrak{J}$.

Now if $N' = n$, we have from the definition of ϕ'

$$\phi - \phi' = P[N \leqq n \mid \mathbf{x}] E[\phi \mid N \leqq n, \mathbf{x}] + P[N > n \mid \mathbf{x}] E[\phi \mid N > n, \mathbf{x}] - \phi'$$
$$= P[N > n \mid \mathbf{x}](E[\phi \mid N > n, \mathbf{x}] - \phi').$$

Thus $|\phi - \phi'| \leqq P[N > n \mid \mathbf{x}]$ if $N' = n$. But $N' = n$ implies $P[N > n \mid \mathbf{x}] < \epsilon$, for all n. This completes the proof of the lemma.

PROOF OF THEOREM 2.1. If condition (2.2) is satisfied, $\mathfrak{I}$ contains a test (N, ϕ) such that

$$\alpha = \sup_{G \varepsilon \mathcal{G}} E_G \phi < \inf_{H \varepsilon \mathcal{H}} E_H \phi = \beta.$$

By the preceding lemma we may and shall assume that N is non-randomized.

Let ϵ be any positive number. The theorem will be proved by showing that there is a non-randomized test (N', ϕ') in $\mathfrak{I}$ such that

$$\inf_{H \varepsilon \mathcal{H}} E_H \phi' - \sup_{G \varepsilon \mathcal{G}} E_G \phi' > 1 - \epsilon. \tag{2.5}$$

Choose a positive integer m which satisfies the inequality

$$\left(\frac{2}{\beta - \alpha}\right)^2 \frac{1}{m} < \frac{\epsilon}{2}.$$

Define the test (N', ϕ') as follows. First apply test (N, ϕ), and denote the resulting values of N and ϕ by N_1 and ϕ_1. Then apply the same test to a new independent sequence of observations and note the values N_2 and ϕ_2 of N and ϕ. Continue in this way until m independent sequences of observations have been taken. The total sample size is $N' = N_1 + \cdots + N_m$. Since N is non-randomized, so is N'. Now put

$$\bar{\phi} = \frac{1}{m} \sum_{i=1}^{m} \phi_i,$$

$$\phi' = \begin{cases} 1 & \text{if } \bar{\phi} > \dfrac{\alpha + \beta}{2} \\ 0 & \text{if } \bar{\phi} \leqq \dfrac{\alpha + \beta}{2}. \end{cases}$$

Thus (N', ϕ') is a non-randomized test.

The chance variables $\phi_1, \cdots, \phi_m$ are independent, and each has the same distribution as ϕ. Hence $E\bar{\phi} = E\,\phi$, and the variance of $\bar{\phi}$ is less than $1/m$.

If $G \varepsilon \mathcal{G}$, then $E_G\phi \leqq \alpha$, so that

$$\begin{aligned} E_G \phi' &= P_G\left[\bar{\phi} - E_G\bar{\phi} > \frac{\alpha + \beta}{2} - E_G \phi\right] \\ &\leqq P_G\left[\bar{\phi} - E_G\bar{\phi} > \frac{\beta - \alpha}{2}\right] \\ &\leqq \frac{1}{m}\left(\frac{2}{\beta - \alpha}\right)^2 \end{aligned}$$

by Chebyshev's inequality. Hence

$$\sup_{G \varepsilon \mathcal{G}} E_G \phi' < \frac{\epsilon}{2}.$$

In a similar way it is seen that

$$\sup_{H \varepsilon \mathfrak{K}} E_H(1 - \phi') < \frac{\epsilon}{2},$$

so that inequality (2.5) is satisfied.

We now show that the test (N', ϕ') is in $\mathfrak{I}$. For $\mathfrak{I} = \mathfrak{I}_0$ and $\mathfrak{I} = \mathfrak{I}_3$ this is obvious. Since for $r > 0$,

$$(N')^r = \left(\sum_{i=1}^{m} N_i\right)^r \leqq (m \max_{i=1,\cdots,m} N_i)^r = m^r \max_{i=1,\cdots,m} (N_i^r) \leqq m^r \sum_{i=1}^{m} N_i^r$$

and each N_i has the same distribution as N, we have $E(N')^r < \infty$ whenever $EN^r < \infty$. This proves the statement for $\mathfrak{I} = \mathfrak{I}_1(r)$ and $\mathfrak{I} = \mathfrak{I}_1$.

Finally, if $Ee^{tN} < \infty$, where $t > 0$, put $t' = t/m$. Since $N_1, \cdots, N_m$ are independent and distributed as N, $Ee^{t'N'} = Ee^{tN} < \infty$.

Thus (N', ϕ') is in $\mathfrak{I}$ in every case. The proof is complete.

It should be noted that if $X_1, X_2, \cdots$ are not independent and identically distributed, the analog of Theorem 2.1 is not true in general.

3. Sufficient conditions for distinguishability. Let $\mathfrak{K}$ be a set of distributions on $\mathfrak{A}$. A distance in $\mathfrak{K}$ is a nonnegative function δ of the pairs (G, H) of distributions in $\mathfrak{K}$ such that $\delta(G, G) = 0$, $\delta(G, H) = \delta(H, G)$, and $\delta(G, H) \leqq \delta(G, K) + \delta(H, K)$, for all G, H, and K in $\mathfrak{K}$. (We do not require that $\delta(G, H) = 0$ imply $G = H$.) We write $\delta(G, \mathfrak{K})$ for $\inf_{H \varepsilon \mathfrak{K}} \delta(G, H)$, and $\delta(\mathfrak{G}, \mathfrak{K})$ for $\inf_{G \varepsilon \mathfrak{G}} \delta(G, \mathfrak{K})$.

Let F_n denote the empiric distribution of the first n members, $X_1, \cdots, X_n$ of a sequence of independent chance variables with the common distribution $F \varepsilon \mathfrak{F}$; that is, $nF_n[A]$ is the number of indices $i \leqq n$ for which $X_i \varepsilon A$. We assume throughout that the set $\mathfrak{K}$ in which a distance δ is defined contains $\mathfrak{F}$ and all empiric distributions.

We shall say that a distance δ is *consistent* in $\mathfrak{F}$ if for every $\epsilon > 0$

$$\lim_{n \to \infty} P_F[\delta(F_n, F) > \epsilon] = 0 \tag{3.1}$$

whenever $F \varepsilon \mathfrak{F}$. The distance δ will be called *uniformly consistent* in $\mathfrak{F}$ if the convergence in (3.1) is uniform for $F \varepsilon \mathfrak{F}$.

In this section we derive sufficient conditions for distinguishability in terms of uniformly consistent distances. We first mention a few examples of such distances. If $\mathfrak{F}$ is the set of all distributions on the k-dimensional Borel field $\mathfrak{A}$, and $\mathfrak{X}$ denotes the k-dimensional Euclidean space, the distance

$$D(G, H) = \sup_{x \varepsilon \mathfrak{X}} | G(x) - H(x) | \tag{3.2}$$

is known to be uniformly consistent in $\mathfrak{F}$ (see, for example, [4]). So is the distance

$$\left(\int_{\mathfrak{X}} | G(x) - H(x) |^r \, dK\right)^{1/r},$$

where $r \geqq 1$ and K is a fixed distribution on $\mathcal{A}$, since it is bounded above by $D(G, H)$. A further example of a uniformly consistent distance is

$$(3.3) \qquad D_\omega(G, H) = D(G_\omega, H_\omega),$$

where G_ω and H_ω are the distributions, according to G and H, of a fixed, real- or vector-valued measurable function ω on $\mathfrak{X}$. If $\mu(F)$ denotes the mean of a one-dimensional distribution F, the distance $| \mu(G) - \mu(H) |$ is uniformly consistent in any class of distributions with bounded variances.

A sufficient condition for finite distinguishability is the following. If the distance δ is uniformly consistent in $\mathcal{G} \cup \mathcal{H}$ and

$$(3.4) \qquad \delta(\mathcal{G}, \mathcal{H}) > 0,$$

then the sets $\mathcal{G}$ and $\mathcal{H}$ are finitely distinguishable.

This can be seen by using the test with $N = n$ fixed and $\phi = 1$ or 0 according as $\delta(F_n, \mathcal{G}) - \delta(F_n, \mathcal{H}) \geqq 0$ or < 0. If $F \varepsilon \mathcal{G}$, then $\delta(F_n, \mathcal{G}) \leqq \delta(F_n, F)$ and $\delta(F_n, \mathcal{H}) \geqq \delta(F, \mathcal{H}) - \delta(F_n, F) \geqq \delta(\mathcal{G}, \mathcal{H}) - \delta(F_n, F)$. Hence $E_F\phi$ does not exceed

$$\sup_{F \varepsilon \mathcal{G} \cup \mathcal{H}} P_F[\delta(F_n, F) \geqq \tfrac{1}{2}\delta(\mathcal{G}, \mathcal{H})].$$

We obtain the same upper bound for $E_F(1 - \phi)$, $F \varepsilon \mathcal{H}$. Our assumptions imply that the bound tends to 0 as $n \to \infty$.

In the proof of the next theorem we shall make use of a test defined as follows. Let δ be a distance, $\{c_i\}$, $i = 1, 2, \cdots$, a sequence of positive numbers, and $\{n_i\}$, $i = 1, 2, \cdots$, an increasing sequence of positive integers. Put

$$\delta_i = \max [\delta(F_{n_i}, \mathcal{G}), \delta(F_{n_i}, \mathcal{H})].$$

Take successive independent samples of sizes $n_1, n_2 - n_1, n_3 - n_2, \cdots$. Continue sampling as long as $\delta_i < c_i$. Stop sampling as soon as $\delta_i \geqq c_i$, and apply the terminal decision function

$$\phi = \begin{cases} 1 \text{ if } \delta(F_{n_i}, \mathcal{G}) \geqq \delta(F_{n_i}, \mathcal{H}) \\ 0 \text{ if } \delta(F_{n_i}, \mathcal{G}) < \delta(F_{n_i}, \mathcal{H}). \end{cases}$$

Thus $N = n_i$, where i is the least integer for which $\delta_i \geqq c_i$.

We shall refer to this test as the test $T(\delta, \{c_i\}, \{n_i\})$.

THEOREM 3.1. (a) *If the distance δ is uniformly consistent in $\mathfrak{F}$, then any two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ for which*

$$(3.6) \qquad \max [\delta(F, \mathcal{G}), \delta(F, \mathcal{H})] > 0 \textit{ if } F \varepsilon \mathfrak{F}$$

are distinguishable $(\mathfrak{F})$.

(b) *If, for every $c > 0$, there exist two positive numbers $A(c)$ and $B(c)$ such that for all integers $n > 0$ and all $F \varepsilon \mathfrak{F}$*

$$(3.7) \qquad P_F[\delta(F_n, F) \geqq c] \leqq A(c)e^{-B(c)n},$$

then any two subsets $\mathcal{G}$ *and* $\mathcal{H}$ *of* $\mathfrak{F}$ *which satisfy* (3.6) *are distinguishable in the class of tests* (N, ϕ) *such that* $E_F e^{tN} < \infty$ *for some* $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.

Proof. Let α be a positive number. Part (a) will be proved by showing that the sequences $\{c_i\}$ and $\{n_i\}$ can be so chosen that the test $(N, \phi) = T(\delta, \{c_i\}, \{n_i\})$ satisfies the conditions

$$E_F\phi \leqq \alpha \text{ if } F \varepsilon \mathcal{G}, \qquad E_F(1 - \phi) \leqq \alpha \text{ if } F \varepsilon \mathcal{H} \tag{3.8}$$

and

$$P_F[N < \infty] = 1 \text{ if } F \varepsilon \mathfrak{F}. \tag{3.9}$$

Let $\{c_i\}$ be a sequence of positive numbers such that

$$\lim_{i\to\infty} c_i = 0. \tag{3.10}$$

Choose the positive numbers $\alpha_1, \alpha_2, \cdots$ so that

$$\sum_{i=1}^{\infty} \alpha_i \leqq \alpha. \tag{3.11}$$

Since δ is uniformly consistent in $\mathfrak{F}$, we can choose the integers $n_1 < n_2 < \cdots$ in such a way that

$$P_F[\delta(F_{n_i}, F) \geqq c_i] \leqq \alpha_i, \qquad i = 1, 2, \cdots \tag{3.12}$$

for all $F \varepsilon \mathfrak{F}$.

If $F \varepsilon \mathcal{G}$,

$$\begin{aligned} E_F\phi &= \sum_{j=1}^{\infty} P_F[\delta_i < c_i \quad \text{for} \quad i < j, \delta_j \geqq c_j, \delta(F_{n_j}, \mathcal{G}) \geqq \delta(F_{n_j}, \mathcal{H})] \\ &\leqq \sum_{j=1}^{\infty} P_F[\delta(F_{n_j}, \mathcal{G}) \geqq c_j] \\ &\leqq \sum_{j=1}^{\infty} P_F[\delta(F_{n_j}, F) \geqq c_j]. \end{aligned}$$

It now follows from (3.12) and (3.11) that $E_F\phi \leqq \alpha$ if $F \varepsilon \mathcal{G}$. In a similar way it is seen that $E_F(1 - \phi) \leqq \alpha$ if $F \varepsilon \mathcal{H}$. Thus the conditions (3.8) are satisfied.

The terminal sample size N takes on the values $n_1, n_2 \cdots$, and we have

$$P_F[N > n_j] = P_F[\delta_i < c_i, i = 1, \cdots, j] \leqq P_F[\delta_j < c_j].$$

By the triangle inequality,

$$\delta_j \geqq \delta^* - \delta(F_{n_j}, F),$$

where

$$\delta^* = \max[\delta(F, \mathcal{G}), \delta(F, \mathcal{H})].$$

By assumption, $\delta^* > 0$ for all $F \varepsilon \mathfrak{F}$.

Hence if $F \varepsilon \mathfrak{F}$,

$$P_F[N > n_j] \leqq P_F[\delta(F_{n_j}, F) > \delta^* - c_j]. \tag{3.13}$$

Since $c_j \to 0$, we have $\delta^* - c_j > c_j$ for j sufficiently large, and then the right side of (3.13) is $\leqq \alpha_j$. By (3.11), $\alpha_j \to 0$ as $j \to \infty$. Thus $P_F[N > n_j] \to 0$ as $j \to \infty$, which implies (3.9). This completes the proof of part (a).

Now suppose that the assumption of part (b) is satisfied. The sequences $\{c_i\}$ and $\{n_i\}$ can be so chosen that, in addition to $\lim c_i = 0$ and $n_i < n_{i+1}$,

$$\liminf_{i\to\infty} i^{-1}(2n_i - n_{i+1}) > 0 \tag{3.14}$$

and

$$\sum_{i=1}^{\infty} A(c_i)e^{-B(c_i)n_i} \leqq \alpha. \tag{3.15}$$

(For instance, put $M(c) = \max [A(c), 1/B(c)]$; choose $c_1, c_2, \cdots$ so that $c_i > 0$, $\lim c_i = 0$ and $M(c_i) \leqq mi^{1/2}$, $i = 1, 2, \cdots$, with a suitable number $m > 0$; and put $n_i = ni$, where n is so large that

$$\sum_{i=1}^{\infty} mi^{1/2}e^{-nm^{-1}i^{1/2}} \leqq \alpha).$$

The inequalities (3.7) and (3.15) imply that conditions (3.11) and (3.12) are fulfilled. Hence the conditions (3.8) are satisfied.

For a fixed $F \varepsilon \mathfrak{F}$, choose the integer h so that $c_i \leqq \delta^*/2$ for $i > h$. Then for $i > h$, due to (3.13) and (3.7),

$$P_F[N > n_i] \leqq P_F[\delta(F_{n_i}, F) > \delta^*/2] \leqq ae^{-bn_i},$$

where $a = A(\delta^*/2)$ and $b = B(\delta^*/2)$ are positive numbers.

Now for any real t,

$$E_F e^{tN} = \sum_{j=1}^{\infty} e^{tn_j}P_F[N = n_j]$$

$$\leqq e^{tn_1} + \sum_{i=1}^{\infty} e^{tn_{i+1}}P_F[N > n_i].$$

Thus $E_F e^{tN} < \infty$ if the series

$$\sum_i e^{tn_{i+1}-bn_i}$$

converges. If $t \leqq b/2$,

$$tn_{i+1} - bn_i \leqq -\frac{b}{2}(2n_i - n_{i+1}),$$

so that the series converges due to (3.14). The proof is complete.

The assumption of Theorem 3.1, part (b) is satisfied if $\mathfrak{F}$ is any set of k-dimen-

sional distributions ($k \geqq 1$) and $\delta = D$, the distance defined by (3.2). This is implied by the following theorem due to Kiefer and one of the authors [4]: For every integer $k \geqq 1$ there exist two positive numbers a and b such that for all $c > 0$, all integers $n > 0$, and all k-dimensional distributions F

$$P_F[D(F_n, F) \geqq c] \leqq ae^{-bc^2n}. \tag{3.16}$$

(For $k = 1$ the inequality (3.16), with $b = 2$, was proved by Dvoretzky, Kiefer and one of the authors [2].) Hence we can state the following corollary.

COROLLARY 3.1. *If $\mathfrak{F}$ is any set of k-dimensional distributions ($k \geqq 1$), then any two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ for which*

$$\max [D(F, \mathcal{G}), D(F, \mathcal{H})] > 0 \text{ if } F \varepsilon \mathfrak{F}$$

are distinguishable in the class of tests (N, ϕ) such that $E_F e^{tN} < \infty$ for some $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.

4. Necessary conditions for distinguishability. Let P and Q be two distributions on a σ-field $\mathcal{B}$ of subsets of an arbitrary space $\mathcal{Y}$, and let Ψ be the class of all measurable functions on $\mathcal{Y}$ with values ranging from 0 to 1. We denote by d the distance defined by

$$d(P, Q) = \sup_{\psi \varepsilon \Psi} | E_P \psi - E_Q \psi |. \tag{4.1}$$

We note some alternative expressions for d. Let ν be any σ-finite measure with respect to which P and Q are absolutely continuous (for instance, $\nu = P + Q$), and denote by p and q densities (Radon-Nikodym derivatives) of P and Q with respect to ν. Then

$$d(P, Q) = \int_{\{p>q\}} (p - q)\, d\nu = \frac{1}{2} \int | p - q |\, d\nu = 1 - \int \min (p, q)\, d\nu. \tag{4.2}$$

(Here and in what follows, an integral whose domain of integration is not indicated is extended over the entire space.) Also

$$d(P, Q) = \sup_{B \varepsilon \mathcal{B}} | P[B] - Q[B] |. \tag{4.3}$$

For any distribution G on $\mathcal{A}$ we denote by $G^{(n)}$ the distribution of n independent chance variables each of which has the distribution G. We write $\mathcal{G}^{(n)}$ for the set of all $G^{(n)}$ such that $G \varepsilon \mathcal{G}$.

It is easily seen from (4.1) that

$$d(G, H) \leqq d(G^{(n)}, H^{(n)}) \leqq d(G^{(n+1)}, H^{(n+1)}), \qquad n = 1, 2, \cdots \tag{4.4}$$

and from the last expression in (4.2), using the inequality $\min (ab, cd) \geqq \min (a, c) \min (b, d)$, where a, b, c, d are all positive, that

$$d(G^{(n)}, H^{(n)}) \leqq 1 - (1 - d(G, H))^n \leqq n\, d(G, H). \tag{4.5}$$

(See also Kruskal [6], p. 29.)

The convex hull, $C\mathcal{P}$, of a set $\mathcal{P}$ of distributions on a common σ-field is defined

as the set of all distributions $\lambda_1 P_1 + \cdots + \lambda_r P_r$, where r is any positive integer, $P_1, \cdots, P_r$ are in $\mathcal{P}$, and $\lambda_1, \cdots, \lambda_r$ are positive numbers whose sum is 1.

In order that two sets $\mathcal{G}$ and $\mathcal{H}$ be finitely distinguishable it is necessary that

$$d(C\mathcal{G}^{(n)}, C\mathcal{H}^{(n)}) > 0 \tag{4.6}$$

for some n or, equivalently,

$$\lim_{n \to \infty} d(C\mathcal{G}^{(n)}, C\mathcal{H}^{(n)}) = 1. \tag{4.7}$$

This is known and follows easily from the definition (4.1) and Theorem 2.1.

If the set $\mathcal{G} \cup \mathcal{H}$ is dominated, that is to say, if the distributions in $\mathcal{G} \cup \mathcal{H}$ are absolutely continuous with respect to a fixed σ-finite measure, then condition (4.7) is also sufficient for $\mathcal{G}$ and $\mathcal{H}$ to be finitely distinguishable. This is contained in Theorem 6 of Kraft [7] and follows from a theorem of LeCam (Theorem 5 of Kraft [7]) which is equivalent to the statement that if the set $\mathcal{P}_1 \cup \mathcal{P}_2$ is dominated, then

$$\max_{\phi \varepsilon \Phi} \inf_{P_1 \varepsilon \mathcal{P}_1, P_2 \varepsilon \mathcal{P}_2} (E_{P_2}\phi - E_{P_1}\phi) = d(C\mathcal{P}_1, C\mathcal{P}_2), \tag{4.8}$$

where Φ denotes the set of all measurable functions ϕ such that $0 \leqq \phi \leqq 1$.

If condition (4.6) is satisfied, then

$$d(\mathcal{G}, \mathcal{H}) > 0. \tag{4.9}$$

In fact, $d(C\mathcal{G}^{(n)}, C\mathcal{H}^{(n)}) \leqq d(\mathcal{G}^{(n)}, \mathcal{H}^{(n)}) \leqq n\, d(\mathcal{G}, \mathcal{H})$ by (4.5). This weaker but much simpler necessary condition for finite distinguishability will be shown in section 5 to be also sufficient under certain assumptions.

To obtain necessary conditions for non-finite distinguishability we first prove the following lemma.

LEMMA 4.1. *If*

$$d(F_0^{(n)}, C\mathcal{G}^{(n)}) = d(F_0^{(n)}, C\mathcal{H}^{(n)}) = 0 \tag{4.10}$$

for all n, then the sets $\mathcal{G}$ and $\mathcal{H}$ are indistinguishable in the class of tests (N, ϕ) with $P_{F_0}[N < \infty] = 1$.

PROOF. Let (N, ϕ) be any test such that $P_{F_0}[N < \infty] = 1$. Define $\phi_n = \phi$ if $N \leqq n$, $\phi_n = 0$ if $N > n$. Thus ϕ_n is a function of $x_1, \cdots, x_n$ only, and $\phi_n \leqq \phi$. Let K be a member of $C\mathcal{G}^{(n)}$, so that $K = \lambda_1 G_r^{(n)} + \cdots + \lambda_r G_r^{(n)}$, $G_i \varepsilon \mathcal{G}$, $\lambda_i > 0$, $\Sigma\lambda_i = 1$. Then[5]

$$E_K \phi_n = \Sigma\lambda_i E_{G_i}\phi_n \leqq \Sigma\lambda_i E_{G_i}\phi \leqq \sup_{G \varepsilon \mathcal{G}} E_G \phi.$$

Hence

$$E_{F_0}\phi_n - \sup_{G \varepsilon \mathcal{G}} E_G \phi \leqq E_{F_0}\phi_n - E_K \phi_n \leqq d(F_0^{(n)}, K)$$

[5] Here $E_K\phi_n$ denotes the expected value of ϕ_n when the joint distribution of $(X_1, \cdots, X_n)$ is K. We keep the notation $E_G\phi_n$ when $X_1, \cdots, X_n$ are independent and each X_i has the distribution G.

for all $K \varepsilon C\mathcal{G}^{(n)}$. Therefore

$$E_{F_0}\phi_n - \sup_{G \varepsilon \mathcal{G}} E_G \phi \leqq d(F_0^{(n)}, C\mathcal{G}^{(n)}) = 0.$$

Since $P_{F_0}(N > n) \to 0$ as $n \to \infty$, $E_{F_0}\phi_n$ converges to $E_{F_0}\phi$. Hence

$$E_{F_0}\phi \leqq \sup_{G \varepsilon \mathcal{G}} E_G \phi. \tag{4.11}$$

In a similar way, if we use, instead of ϕ_n, the function $\phi_n' = \phi$ if $N \leqq n$, $\phi_n' = 1$ if $N > n$, we find that

$$E_{F_0}\phi \geqq \inf_{H \varepsilon \mathcal{H}} E_H \phi. \tag{4.12}$$

Inequalities (4.11) and (4.12) imply the Lemma.

THEOREM 4.1. *In order that the sets $\mathcal{G}$ and $\mathcal{H}$ be distinguishable ($\mathfrak{F}$) it is necessary that*

$$\max\,[d(F^{(n)}, C\mathcal{G}^{(n)}), d(F^{(n)}, C\mathcal{H}^{(n)})] > 0 \tag{4.13}$$

for some n if $F \varepsilon \mathfrak{F}$ and hence that

$$\max\,[d(F, \mathcal{G}), d(F, \mathcal{H})] > 0 \text{ if } F \varepsilon \mathfrak{F}. \tag{4.14}$$

PROOF. The necessity of (4.14) follows immediately from Lemma 4.1. That (4.13) implies (4.14) follows from inequality (4.5).

That the condition (4.13) can be violated when inequality (4.14) is satisfied can be seen from an example given by Kraft ([7], p. 132) to show the non-equivalence of two conditions equivalent to (4.6) and (4.9). Nevertheless the simple necessary condition (4.14) is also sufficient under certain restrictions on the set of distributions, as will be seen in section 5.

We conclude this section by showing that a known necessary condition for distinguishability is implied by condition (4.14) of Theorem 4.1.

For any two distributions F and G on $\mathcal{A}$ and any set $\mathcal{G}$ of distributions on $\mathcal{A}$ define

$$\tau(F, G) = \int f \log (f/g)\, d\nu, \qquad \tau(F, \mathcal{G}) = \inf_{G \varepsilon \mathcal{G}} \tau(F, G),$$

where ν denotes a σ-finite measure with respect to which F and G are absolutely continuous, with densities f and g. Note that $0 \leqq \tau\,(F, G) \leqq \infty$. It has been shown in [3] that if $\tau(F, \mathcal{G}) = 0$, then F and $\mathcal{G}$ are indistinguishable in the class of tests with $E_F N < \infty$. Now

$$-\tfrac{1}{2}\tau(F, G) = \int f \log (g/f)^{1/2}\, d\nu$$

$$\leqq \log \int f(g/f)^{1/2}\, d\nu = \log \int (fg)^{1/2}\, d\nu$$

and (see Kraft [7], Lemma 1)

$$d^2(F, G) \leqq 1 - \left\{\int (fg)^{1/2}\, d\nu\right\}^2.$$

Hence $\tau(F, \mathcal{G}) = 0$ implies $d(F, \mathcal{G}) = 0$. Thus, by Lemma 4.1 (with $F_0 = F$ and $\mathcal{H}$ consisting only of F) F and $\mathcal{G}$ are even indistinguishable in the class of tests with $P_F[N < \infty] = 1$. It is easy to construct examples where $d(F, \mathcal{G}) = 0$ and $\tau(F, \mathcal{G}) > 0$, so that condition (4.14) is actually better than the corresponding condition with d replaced by τ.

5. Necessary and sufficient conditions for distinguishability. In this section we shall show that the necessary conditions of section 4 are also sufficient for distinguishability under certain restrictions on the sets of distributions. Most of our results will be such that if the necessary condition is satisfied, the sets are not only distinguishable ($\mathcal{F}$), but even distinguishable in a stronger sense.

If $\mathcal{G}$ consists of a single distribution G, then, by Theorem 4.1, $\mathcal{G}$ and $\mathcal{H}$ are distinguishable ($\mathcal{G} \cup \mathcal{H}$) only if $d(G^{(n)}, C\mathcal{H}^{(n)}) > 0$ for some n. If $\mathcal{H}$ is dominated, this condition is sufficient for $\mathcal{G}$ and $\mathcal{H}$ to be finitely distinguishable, by the Le Cam-Kraft theorem mentioned in section 4. More generally, we can state the following.

If $\mathcal{G}$ is finite and $\mathcal{H}$ is dominated, then $\mathcal{G}$ and $\mathcal{H}$ are either finitely distinguishable or are indistinguishable ($\mathcal{G} \cup \mathcal{H}$), depending on whether the condition

$$\max [d(F^{(n)}, C\mathcal{G}^{(n)}), d(F^{(n)}, C\mathcal{H}^{(n)})] > 0 \tag{5.1}$$

for some n if $F \varepsilon \mathcal{G} \cup \mathcal{H}$ is or is not satisfied. Condition (5.1) is equivalent to

$$d(G^{(n)}, C\mathcal{H}^{(n)}) > 0 \tag{5.2}$$

for some n if $G \varepsilon \mathcal{G}$.

That condition (5.1) is necessary for $\mathcal{G}$ and $\mathcal{H}$ to be distinguishable ($\mathcal{G} \cup \mathcal{H}$) follows from Theorem 4.1. On the other hand, if (5.1) is satisfied, so is (5.2). Hence if the distributions in $\mathcal{G}$ are denoted by $G_1, \cdots, G_r$, then, by Le Cam's theorem, G_i and $\mathcal{H}$ are finitely distinguishable, for each i. Thus, given $\epsilon > 0$, there exists an integer n and tests (n, ϕ_i) such that $E_{G_i}\phi_i < \epsilon$ and

$$\sup_{H \varepsilon \mathcal{H}} E_H(1 - \phi_i) < \epsilon, i = 1, \cdots, r.$$

Put $\phi = \phi_1\phi_2 \cdots \phi_r$. Then $\phi \leqq \phi_i$ and $1 - \phi \leqq \sum_{i=1}^{r} (1 - \phi_i)$. Hence $E_{G_i}\phi < \epsilon$ for all i and $E_H (1 - \phi) < r\epsilon$ if $H \varepsilon \mathcal{H}$. Therefore $\mathcal{G}$ and $\mathcal{H}$ are finitely distinguishable, and condition (5.2) is equivalent to (5.1).

If both $\mathcal{G}$ and $\mathcal{H}$ are countably infinite sets, it is no longer true that $\mathcal{G}$ and $\mathcal{H}$ are either finitely distinguishable or indistinguishable. To see this, let $\mathcal{G} = \{G_i\}$ and $\mathcal{H} = \{H_i\}$, $i = 1, 2, \cdots$, where G_i and H_i are univariate normal distributions with respective means a and b ($a \neq b$) and common variance σ_i^2, such that $\lim \sigma_i^2 = \infty$. It follows from a result of Stein [8] (or from Theorem 3.1) that $\mathcal{G}$ and $\mathcal{H}$ are distinguishable ($\mathfrak{N}$), where $\mathfrak{N}$ denotes the class of univariate normal distributions. But one readily verifies that $\lim_i d(G_i, H_i) = 0$, so that the sets are finitely indistinguishable.

In what follows it will be shown that the simple condition

$$\max [d(F, \mathcal{G}), d(F, \mathcal{H})] > 0 \text{ if } F \varepsilon \mathcal{F} \tag{5.3}$$

which, by Theorem 4.1, is necessary for $\mathcal{G}$ and $\mathcal{H}$ to be distinguishable ($\mathcal{F}$), is also

sufficient under rather general assumptions. Under somewhat more stringent assumptions the necessary condition $d(\mathcal{G}, \mathcal{H}) > 0$ for finite distinguishability (see (4.9)) will also be shown to be sufficient.

A comparison of the results of sections 3 and 4 shows that if δ is any uniformly consistent distance in a set $\mathfrak{F}$, then $d(\mathcal{G}, \mathcal{H}) = 0$ implies $\delta(\mathcal{G}, \mathcal{H}) = 0$ whenever $\mathcal{G} \subset \mathfrak{F}$ and $\mathcal{H} \subset \mathfrak{F}$. Theorems 3.1 and 4.1 also show that if the set $\mathfrak{F}$ has the property that there exists a uniformly consistent δ such that, for all $F \varepsilon \mathfrak{F}$ and all $\mathcal{G} \subset \mathfrak{F}$, $\delta(F, \mathcal{G}) = 0$ implies (and hence is equivalent to) $d(F, \mathcal{G}) = 0$, then the necessary condition (4.14) is also sufficient for two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ to be distinguishable ($\mathfrak{F}$). Similarly, if for all $\mathcal{G} \subset \mathfrak{F}$ and all $\mathcal{H} \subset \mathfrak{F}$, $\delta(\mathcal{G}, \mathcal{H}) = 0$ implies $d(\mathcal{G}, \mathcal{H}) = 0$, then any two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ are finitely distinguishable if and only if $d(\mathcal{G}, \mathcal{H}) > 0$.

We first consider conditions which ensure that $D(F, \mathcal{G}) = 0$ implies $d(F, \mathcal{G}) = 0$. Let F and G be two k-dimensional distributions and ϵ a nonnegative number. Suppose that there is an integer J with the following property. There exist J non-overlapping k-dimensional intervals $I_1, \quad , I_J$ such that (i) $F - G$ is monotone[6] in each I_j, and (ii) if V denotes the complement of $\bigcup_{j=1}^{J} I_j$, then $\min(F[V], G[V]) \leqq \epsilon$. Write $J(F, G; \epsilon)$ for the least integer J having this property. If such a finite J does not exist, define $J(F, G; \epsilon) = \infty$.

Note that if $F - G$ is monotone in a set C, the difference of the densities, $f - g$, is of constant sign in C except in a subset of probability 0 according to both F and G.

LEMMA 5.1. *If F and G are two k-dimensional distributions,*

$$d(F, G) \leqq 2^k J(F, G; \epsilon)\, D(F, G) + \epsilon. \tag{5.4}$$

PROOF. We may assume that $J = J(F, G, \epsilon)$ is finite. Then there exist J non-overlapping intervals $I_1, \cdots, I_J$ which satisfy the conditions (i) and (ii). We have

$$2d(F, G) = \sum_{j=1}^{J} \int_{I_j} | f - g | \, d\nu + \int_V | f - g | \, d\nu.$$

Now

$$\int_V | f - g | \, d\nu \leqq \int_V (f + g) \, d\nu = 2 \int_V f d\nu + \int_V (g - f) \, d\nu$$

$$= 2F[V] + \sum_{j=1}^{J} \int_{I_j} (f - g) \, d\nu \leqq 2F[V] + \sum_{j=1}^{J} \left| \int_{I_j} (f - g) \, d\nu \right| .$$

Also,

$$\int_{I_j} | f - g | \, d\nu = \left| \int_{I_j} (f - g) \, d\nu \right| \leqq 2^k D\,(F, G).$$

[6] An additive function L on $\mathfrak{R}$ is monotone in a set C is either $L \leqq [A]\ L\ [B]$ whenever $A \subset B \subset C$ or $L\,[A] \geqq L\,[B]$ whenever $A \subset B \subset C$.

Hence

$$2d(F, G) \leqq 2J \cdot 2^k D(F, G) + 2F[V].$$

By symmetry, the term $2F[V]$ can be replaced by $2G[V]$, and hence also by 2ϵ. This implies (5.4).

THEOREM 5.1. *Let $\mathfrak{F}$ be a set of k-dimensional distributions, $k \geqq 1$.* (a) *If*

$$\sup_{G \varepsilon \mathfrak{F}} J(F, G; \epsilon) < \infty \tag{5.5}$$

for all $F \varepsilon \mathfrak{F}$ and all $\epsilon > 0$, then two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ are distinguishable ($\mathfrak{F}$) if and only if

$$\max [d(F, \mathcal{G}), d(F, \mathcal{H})] > 0 \tag{5.6}$$

for all $F \varepsilon \mathfrak{F}$. Moreover, if condition (5.6) *is satisfied, then $\mathcal{G}$ and $\mathcal{H}$ are distinguishable in the class of tests (N, ϕ) such that $E_F e^{tN} < \infty$ for some $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.* (b) *If*

$$\sup_{F \varepsilon \mathfrak{F}, G \varepsilon \mathfrak{F}} J(F, G; \epsilon) < \infty \tag{5.7}$$

for all $\epsilon > 0$, then two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ are finitely distinguishable if and only if

$$d(\mathcal{G}, \mathcal{H}) > 0. \tag{5.8}$$

PROOF. The necessity of conditions (5.6) and (5.8) has been proved in section 4. If condition (5.5) is satisfied, then, by Lemma 5.1, $D(F, \mathcal{G}) = 0$ implies $d(F, \mathcal{G}) = 0$ for all $F \varepsilon \mathfrak{F}$ and all $\mathcal{G} \subset \mathfrak{F}$. Hence if (5.6) is satisfied, the assumption of Corollary 3.1 is fulfilled, which implies part (a). The proof of part (b) is similar, referring to (3.4) with $\delta = D$.

The assumption of Theorem 5.1, part (b) (and hence that of part (a)) is satisfied for most parametric sets of *univariate* distributions which are commonly used as models in statistics. In such sets $\mathfrak{F}$ the minimum number of intervals in which $f - g$ is of constant sign is usually bounded, and then even $\sup_{F \varepsilon \mathfrak{F}, G \varepsilon \mathfrak{F}} J(F, G; 0)$ is finite. For example, if F and G are any two univariate normal distributions, then $J(F, G; 0) \leqq 3$. This is also true if the singular normal distributions (with zero variance) are included.

The assumption of part (a) is satisfied if $\mathfrak{F}$ is any subclass of the class of all distributions on the subsets of a fixed countable set S. Since the points of S can be arranged in a sequence, we may assume that S is the set of the positive integers. If $F \varepsilon \mathfrak{F}$ and $\epsilon > 0$, choose the integer M so that $F[x > M] < \epsilon$. Since we can choose M intervals each of which contains exactly one positive integer $\leqq M$, we have $J(F, G; \epsilon) \leqq M$ for all $G \varepsilon \mathfrak{F}$, so that condition (5.5) is satisfied.

Actual statistical observations are either integer-valued or integer multiples of a fixed unit of measurement. In this sense it can be said that the assumption of part (a) is satisfied for all classes of distributions which actually occur in statistics.

If $\mathcal{G}$ and $\mathcal{H}$ are two arbitrary sets of distributions over a fixed countable set,

then $\mathcal{G}$ and $\mathcal{H}$ can be *finitely* indistinguishable even when $d(\mathcal{G}, \mathcal{H}) > 0$. This is shown by the following example. For $r = 1, 2, \cdots$ and $k = 1, \cdots, r$ define the sets

$$A_r = \{i2^{-r} \mid i = 1, 2, \cdots, 2^r\},$$
$$A_{r,k} = \{(j2^{r-k+1} + i)2^{-r} \mid i = 1, 2, \cdots, 2^{r-k}; j = 0, 1, \cdots, 2^{k-1} - 1\}.$$

Let G_r and $H_{r,k}$ be the discrete distributions whose elementary probability functions are

$$g_r(x) = 2^{-r}\chi(x; A_r), \qquad h_{r,k}(x) = 2^{-r+1}\chi(x; A_{r,k}),$$

where $\chi(x; A) = 1$ or 0 according as $x \varepsilon A$ or $x \not\varepsilon A$. Let $\mathcal{G} = \{G_r\}, r = 1, 2, \cdots$, and $\mathcal{H} = \{H_{r,k}\}, k = 1, \cdots, r; r = 1, 2, \cdots$. The reader can verify that

$$d(G_r, H_{s,k}) \geq \tfrac{1}{2}$$

for all r, s, and k, so that $d(\mathcal{G}, \mathcal{H}) > 0$.

Now denote by $G_r^{(n)}$ and $H_{r,k}^{(n)}$ the distributions of n independent chance variables each of which has the distribution G_r and $H_{r,k}$, respectively, and by $g_r^{(n)}$ and $h_{r,k}^{(n)}$ their elementary probability functions Let $H_r^{(n)}$ denote the distribution in $C\mathcal{H}^{(n)}$ whose elementary probability function is

$$h_r^{(n)} = r^{-1} \sum_{k=1}^{r} h_{r,k}^{(n)}.$$

Writing $g_r^{(n)}$, g_r, etc. for the chance variables $g_r^{(n)}(X_1, \cdots, X_n)$, $g_r(X_1)$, etc., and E for the expected value when the distribution of X_i is G_r, we have

$$2\, d(G_r^{(n)}, H_r^{(n)}) = E \mid (h_r^{(n)}/g_r^{(n)}) - 1 \mid \leq (E[(h_r^{(n)}/g_r^{(n)}) - 1]^2)^{1/2}$$
$$= (E(h_r^{(n)}/g_r^{(n)})^2 - 1)^{1/2}$$

We calculate

$$E(h_r^{(n)}/g_r^{(n)})^2 = r^{-2} \sum_{j=1}^{r} \sum_{k=1}^{r} (E(h_{rj}\, h_{rk}/g_r^2))^n = 1 + (2^n - 1)r^{-1}.$$

It follows that $\lim_{r\to\infty} d(G_r^{(n)}, H_r^{(n)}) = 0$ for every n. Therefore $d(\mathcal{G}^{(n)}, C\mathcal{H}^{(n)}) = 0$ for all n, so that the sets $\mathcal{G}$ and $\mathcal{H}$ are finitely indistinguishable. Note, however, that since $d(\mathcal{G}, \mathcal{H}) > 0$, the sets are distinguishable in the sense of part (a) of Theorem 5.1 with $\mathcal{F} = \mathcal{G} \cup \mathcal{H}$ and, more generally, with $\mathcal{F}$ denoting any class of distributions on the subsets of $\bigcup_{r=1}^{\infty} A_r$ such that condition (5.6) is satisfied.

We shall see that all conclusions of Theorem 5.1 are true also for arbitrary sets of k-dimensional normal distributions, for any $k \geq 1$. However, for $k > 1$ this cannot be deduced from Theorem 5.1 since the multidimensional D distance does not have the properties required by the theorem. It can be shown that if $\mathcal{F}$ is any set of non-singular bivariate normal distributions, the assumption of part (a) is satisfied. But for arbitrary sets of bivariate (possibly singular) normal distributions, $D(F, \mathcal{G}) = 0$ does not imply $d(F, \mathcal{G}) = 0$. (For instance, if F_c

denotes the bivariate normal distribution with means $(c, -c)$, unit variances and correlation coefficient 1, and $\mathcal{G} = \{F_c \mid c > 0\}$, then $D(F_0, \mathcal{G}) = 0$ but $d(F_0, \mathcal{G}) = 1$.) Moreover, $D(\mathcal{G}, \mathcal{H}) = 0$ does not imply $d(\mathcal{G}, \mathcal{H}) = 0$ even for sets of non-singular bivariate normal distributions. (Thus if G_c denotes the bivariate normal distribution with means $(c, -c)$, unit variances, and correlation coefficient $(1 + c^2)^{-1}$, if $\mathcal{G} = \{G_c \mid c < 0\}$ and $\mathcal{H} = \{G_c \mid c > 0\}$, then $D(\mathcal{G}, \mathcal{H}) = 0$ but $d(\mathcal{G}, \mathcal{H}) > 0$.)

For a fixed $k \geqq 1$ let $\mathfrak{N}$ denote the set of all k-dimensional normal distributions. To prove the statement at the beginning of the preceding paragraph it is sufficient to display a distance δ such that $\delta(\mathcal{G}, \mathcal{H}) = 0$ implies $d(\mathcal{G}, \mathcal{H}) = 0$ whenever $\mathcal{G} \subset \mathfrak{N}$ and $\mathcal{H} \subset \mathfrak{N}$, and δ satisfies assumption (3.7) of Theorem 3.1 with $\mathfrak{F} = \mathfrak{N}$. We shall show this to be true for the distance δ^* defined as follows.

For any k-dimensional distribution F with finite moments of the second order define $\theta(F) = (\mu(F), \Sigma(F))$, where $\mu(F)$ denotes the vector of the means and $\Sigma(F)$ the covariance matrix of F. Denote by Θ the range of $\theta(F)$. Define the function $d^*(\theta_1, \theta_2)$, $\theta_1, \theta_2 \in \Theta$ by

$$d^*(\theta_1, \theta_2) = d(F_1, F_2) \text{ if } F_i \in \mathfrak{N} \quad \text{and} \quad \theta(F_i) = \theta_i, \qquad i = 1, 2.$$

Now define δ^* by

$$\delta^*(F_1, F_2) = d^*(\theta(F_1), \theta(F_2))$$

for any two k-dimensional distributions F_1 and F_2 with finite moments of the second order.

The function δ^* is a distance[7] in the set of distributions for which it is defined. Obviously $\delta^*(\mathcal{G}, \mathcal{H}) = 0$ if and only if $d(\mathcal{G}, \mathcal{H}) = 0$ for $\mathcal{G} \subset \mathfrak{N}$ and $\mathcal{H} \subset \mathfrak{N}$.

Now let F_n be the empiric distribution of n independent chance variables $X_1, \cdots, X_n$, each of which has the distribution $F \in \mathfrak{N}$. Put $\theta(F) = \theta = (\mu, \Sigma)$ and $\theta(F_n) = \hat{\theta} = (\hat{\mu}, \hat{\Sigma})$. Thus $\hat{\mu}$ is the sample mean vector and $\hat{\Sigma}$ the sample covariance matrix. We have

$$\delta^*(F_n, F) = d^*(\hat{\theta}, \theta).$$

It follows from the definition of d^* that the distribution of $d^*(\hat{\theta}, \theta)$ does not change if each X_i is subjected to the same non-singular linear transformation. Hence the distribution of $d^*(\hat{\theta}, \theta)$ depends only on the rank r of Σ. If $r = k$, we may assume that $\theta = (0, I) = \theta_0$ (say), where 0 denotes the zero vector with k components and I the $k \times k$ unit matrix. If $1 \leqq r < k$, the distribution of $d^*(\hat{\theta}, \theta)$ is the same, only with k replaced by r. If $r = 0$, $d^*(\hat{\theta}, \theta) = 0$ with probability one. Thus we may confine ourselves to the case $r = k$, $\theta = \theta_0$. We have only to show that for every $c > 0$ there exist numbers $A(c)$ and $B(c)$ such that for all integers $n > 0$,

$$P[d^*(\hat{\theta}, \theta_0) > c] < A(c)e^{-B(c)n}. \tag{5.9}$$

[7] Recall that $\delta^*(F_1, F_2) = 0$ need not imply $F_1 = F_2$.

Now the function $d^*(\theta, \theta_0)$ is continuous at $\theta = \theta_0$ in the usual sense. Hence it is easily seen that (5.9) is satisfied if for every $\epsilon > 0$ the probability of each of the inequalities

$$|\hat{\mu}_i| > \epsilon, \qquad |\hat{\sigma}_{ii} - 1| > \epsilon, \qquad |\hat{\rho}_{ij}| > \epsilon, \qquad i \neq j,\ i, j = 1, \cdots, k,$$

where $\hat{\rho}_{ij} = \hat{\sigma}_{ij}(\hat{\sigma}_{ii}\hat{\sigma}_{jj})^{-1/2}$, and $\hat{\mu}_i$ and $\hat{\sigma}_{ij}$ are the components of $\hat{\mu}$ and $\hat{\Sigma}$, does not exceed a bound of the form $A(\epsilon)\exp(-B(\epsilon)n)$ with $B(\epsilon) > 0$. That the latter is true is seen by considering the well-known distributions of $\hat{\mu}_i$, $\hat{\sigma}_{ii}$, and $\hat{\rho}_{ij}$. This completes the proof.

In the proof we could have equally well used, instead of d, the distance

$$d_1(F, G) = \left\{\int (f^{1/2} - g^{1/2})^2\, d\nu\right\}^{1/2} = 2^{1/2}\{1 - \rho(F, G)\}^{1/2}$$

where ν denotes a measure such that F and G have densities, f and g, with respect to ν, and

$$\rho(F, G) = \int (fg)^{1/2}\, d\nu.$$

For we have (see, for instance, Kraft [7], Lemma 1)

$$1 - \rho(F, G) \leqq d(F, G) \leqq (1 - \rho^2(F, G))^{1/2},$$

so that the distances d and d_1 are equivalent for our purposes.

Define $d_1^*(\theta_1, \theta_2)$ and $\delta_1^*(F, G)$ in terms of d_1 just like d^* and δ^* were defined in terms of d. We shall write $\rho(\theta_1, \theta_2)$ for $\rho(F_1, F_2)$ if $F_i \varepsilon \mathfrak{N}$, $\theta(F_i) = \theta_i$. Thus $d_1^*(\theta_1, \theta_2) = 2^{1/2}(1 - \rho(\theta_1, \theta_2))^{1/2}$. If Σ_1 and Σ_2 are nonsingular,

$$(5.10) \qquad \rho(\theta_1, \theta_2) = |\Sigma_1|^{1/4}|\Sigma_2|^{1/4}\left|\frac{\Sigma_1 + \Sigma_2}{2}\right|^{-1/2} \times \exp\left\{-\frac{1}{4}(\mu_1 - \mu_2)'(\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2)\right\},$$

where μ_1 and μ_2 are regarded as column vectors and the prime denotes the transpose. (Compare Kraft [7], p. 129, where there are some misprints.) If Σ_1 has rank r, $1 \leqq r < k$, then $\rho(\theta_1, \theta_2) = 0$ unless Σ_2 also has rank r and the normal distributions with $\theta = \theta_1$ and $\theta = \theta_2$ assign probability one to the same r-dimensional plane, H; in this case $\rho(\theta_1, \theta_2)$ is equal to an expression like (5.10), with μ_i and Σ_i now denoting the means and covariances, in a common coordinate system, of the corresponding r-dimensional normal distributions on H. If the rank of Σ_1 is 0, then $\rho(\theta_1, \theta_2) = 0$ or 1 according as $\theta_1 \neq \theta_2$ or $\theta_1 = \theta_2$.

If Γ and Δ are subsets of Θ, write $\rho(\theta, \Delta)$ for $\sup_{\theta' \varepsilon \Delta} \rho(\theta, \theta')$ and $\rho(\Gamma, \Delta)$ for $\sup_{\theta \varepsilon \Gamma} \rho(\theta, \Delta)$. If $\mathcal{G} \subset \mathfrak{N}$, define $\theta(\mathcal{G}) = \{\theta(F) \mid F \varepsilon \mathcal{G}\}$.

Expressing the conditions (5.6) and (5.8) of Theorem 5.1 in terms of ρ, we can summarize the foregoing as follows.

THEOREM 5.2. *Let $\mathfrak{N}$ be the set of all k-dimensional normal distributions, $k \geq 1$.*
(a) *If $\mathfrak{F} \subset \mathfrak{N}$, then two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ are distinguishable ($\mathfrak{F}$) if and only if*

$$\min [\rho(\theta(F), \theta(\mathcal{G})), \rho(\theta(F), \theta(\mathcal{H}))] < 1 \tag{5.11}$$

for all $F \varepsilon \mathfrak{F}$. Moreover, if condition (5.11) is satisfied, $\mathcal{G}$ and $\mathcal{H}$ are distinguishable in the class of tests (N, ϕ) such that $E_F e^{tN} < \infty$ for some $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.
(b) *Two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{N}$ are finitely distinguishable if and only if*

$$\rho(\theta(\mathcal{G}), \theta(\mathcal{H})) < 1. \tag{5.12}$$

We observe that condition (5.11) can be expressed in an alternative form. Note that $\rho(\theta_1, \theta_2) = 1$ if and only if $\theta_1 = \theta_2$. If $\theta = (\mu, \Sigma) \varepsilon \Theta$, where Σ is nonsingular, and $\Delta \subset \Theta$, then $\rho(\theta, \Delta) = 1$ if and only if there is a sequence $\{\theta_i\}$ in Δ such that each of the real components of θ_i converges to the corresponding component of θ(in the ordinary sense). If Σ is singular of rank r, the same is true, but with the additional condition that the normal distributions with parameters θ_i and θ assign probability one to the same r-dimensional plane. Thus, for instance, if $\mathfrak{F}$ is a set of non-singular distributions, condition (5.11) is equivalent to the statement that, for every $F \varepsilon \mathfrak{F}$, the Euclidean distance of $\theta(F)$ from $\theta(\mathcal{G})$ or from $\theta(\mathcal{H})$ is positive.

Condition (5.12) does not seem to have an equally simple interpretation.

By way of illustration, let $\mathcal{G}$ and $\mathcal{H}$ denote two sets of univariate normal distributions with positive variances such that $\mu < 0$ if $(\mu, \sigma^2) \varepsilon \theta(\mathcal{G})$ and $\theta(\mathcal{H}) = \{(\mu, \sigma^2) \mid (-\mu, \sigma^2) \varepsilon \theta(\mathcal{G})\}$. Then $\mathcal{G}$ and $\mathcal{H}$ are finitely distinguishable if and only if μ/σ is bounded away from 0 in $\theta(\mathcal{H})$. They are always distinguishable ($\mathcal{G} \cup \mathcal{H}$). If $\mathfrak{F}$ denotes a set of normal distributions with positive variances which contains $\mathcal{G} \cup \mathcal{H}$, then $\mathcal{G}$ and $\mathcal{H}$ are distinguishable ($\mathfrak{F}$) if and only if the distance of every point $(0, \sigma^2) \varepsilon \theta(\mathfrak{F})$ from $\theta(\mathcal{H})$ is positive.

REFERENCES

[1] A. BERGER, "On uniformly consistent tests", *Ann. Math. Stat.*, Vol. 22 (1951), pp. 289–293.

[2] A. DVORETZKY, J. KIEFER, AND J. WOLFOWITZ, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator", *Ann. Math. Stat.*, Vol. 27 (1956), pp. 642–669.

[3] W. HOEFFDING, "The role of assumptions in statistical decisions", *Proc. Third Berkeley Symposium on Math. Stat. and Prob.*, University of California Press (1956), pp. 105–116.

[4] J. KIEFER AND J. WOLFOWITZ, "On the deviations of the empiric distribution of vector chance variables", to appear in *Trans. Amer. Math. Soc.*

[5] A. KOLMOGOROV, *Foundations of the Theory of Probability* (English translation), Chelsea, 1950.

[6] W. H. KRUSKAL, "On the problem of nonnormality in relation to hypothesis testing". (Dittoed)

[7] C. KRAFT, "Some conditions for consistency and uniform consistency of statistical procedures". University of California Publications in Statistics, Vol. 2 (1955), pp. 125–142.

[8] C. STEIN, "A two-sample test for a linear hypothesis whose power is independent of the variance", *Ann. Math. Stat.*, Vol. 16 (1945), pp. 243–258.

THE STRUCTURE OF BIVARIATE DISTRIBUTIONS

BY. H. O. LANCASTER

School of Public Health and Tropical Medicine, Sydney, Australia

1. Introduction. K. Pearson [18] in his study on the association between two chance variables defined a measure, the mean square contingency, $\phi^2 = \chi^2/N$, where χ^2 is that, usually calculated in a contingency table with fixed marginal totals, and N is the size of the sample. In a bivariate joint normal distribution with coefficient of correlation, ρ, Pearson showed that ϕ^2 would have a limiting value if the sample size became indefinitely large, while the subdivisions of the marginal distributions were made increasingly fine. In effect, he was considering a property of the parent joint normal distribution, rather than of a sample drawn from it. He noted that this limiting ϕ^2 was independent of the scale of the marginal variables and was invariant under any bi-unique transformations of the marginal variables of the form, $x \rightarrow x'(x)$, $y \rightarrow y'(y)$. If the distribution was the bivariate joint normal, he showed that $\rho^2 = \phi^2/(1 + \phi^2)$. In some distributions, jointly normal with appropriate choice of the marginal variable, but not so with the variables actually chosen, he took the value of ρ^2 still to have the meaning that an appropriate transformation would yield the variables of the underlying joint normal distribution.

Hirshfeld [8], considering contigency tables with a finite number of discrete values of the variables, sought for transformations of the marginal variables that would yield linear least squares regression lines. He found that these variables maximised the coefficients of correlation

Fisher [3] defined a set of variables on each of the marginal distributions of an $m \times n$ contigency table, such that $x_j = 1$ for an observation falling into the jth class and $x_j = 0$ elsewhere for $j = 1, 2 \cdots m - 1$, and similarly for y_j with $j = 1, 2 \cdots (n - 1)$. His problem was to find a linear form in the x_j, which would have maximum correlation with any linear form in the y_j. For convenience, these linear forms were considered without loss of generality as being normalised. Fisher referred to such a variable and the corresponding correlation as canonical and thus identified them with the canonical variables and correlation of Hotelling [10]. Fisher's theory was amplified by Maung [13] and Williams [25], who considered observational data in the form of a contingency table. We shall see later that in this case, the problem of finding the canonical correlations is equivalent to the determination of the canonical form of a rectangular matrix under pre- and post-multiplication by orthogonal matrices.

It is of interest to extend this type of analysis to the theoretical parent population and to more general classes of bivariate distributions. Lancaster [12] applied the methods of the theory of integral equations to find the canonical correlations and variables in the joint normal distribution and this work leads to a generalisa-

Received September 30, 1957; revised December 10, 1957.

tion of the canonical correlation theory. If the correlation is to have meaning, the canonical variables must have a finite variance, so that each canonical variable can be expressed as an orthonormal linear form in a complete set of orthogonal functions defined on the marginal distribution. The problem is now one in eigenvalue theory. Indeed, it is shown that the canonical correlations are the eigenvalues and the canonical variables on each marginal distribution form a subset, perhaps improper, of a complete set; the canonical variables are, moreover, the eigenfunctions except for a factor. This analysis holds provided the limiting value of Pearson's ϕ^2 is finite. If ϕ^2 is finite, it is further shown that the bivariate distribution can be expanded in an eigenfunction expansion. ϕ^2 is then the sum of the squares of the canonical correlations. The contingency table is then shown to be a special case of the general theory.

Once the canonical form of a bivariate population, that is, the eigenfunction expansion, has been obtained, some further applications of the theory can be made. First, the regressions take a particularly simple form and are confirmed to be the solution of Hirschfeld's problem. Second, given the marginal distributions it is possible to obtain bivariate distributions with prescribed correlations. Third, a goodness of fit test can be devised for the bivariate joint normal distribution, which displays as components of χ^2, the contributions of the regressions of the ith Hermite-Chebishev polynomial in x on the jth polynomial in y. The test is made of the total contributions from those pairs for which $i \neq j$.

2. Pearson's ϕ^2 as the Sum of Squares of the Correlation Coefficients. K. Pearson [18] introduced ϕ^2 as the "mean square contigency" for a bivariate distribution in order to derive a measure of association independent of the sample size, N. He wrote $\phi^2 = \chi^2/N$. Pearson saw that χ^2 (or rather ϕ^2) had a use as a descriptive measure, whereas it is usually thought of as a criterion of goodness of fit, e.g., as in the test due to Pearson [16]. It is convenient to modify Pearson's definition by using the integral sign in the sense of Lebesgue-Stieltjes and adopting the notation of Hellinger [7], which has been justified by Hobson [9].

DEFINITION.

$$\phi^2 = \iint_{-\infty}^{+\infty} [dF(x, y)]^2/[dG(x)\, dH(y)] - 1 \tag{1A}$$

$$= \iint_{-\infty}^{\infty} \Omega^2(x, y)\, dG(x)\, dH(y) - 1 \tag{1B}$$

where

$$\Omega(x, y) = dF(x, y)/[dG(x)\, dH(y)]. \tag{2}$$

$\Omega(x, y)$, and so the integrand of (1A), is to be taken as zero, if the point (x, y) does not correspond to points of increase of both $G(x)$ and $H(y)$. ϕ^2 can evidently be regarded as the limit of the sum $\sum_{i,j} f_{ij}^2/(f_{i.}f_{.j}) - 1$, where f_{ij} is the weight of the bivariate distribution corresponding to marginal sets, A_i and B_j, and where $f_{i.}$ and $f_{.j}$ are the weights of the marginal distributions corresponding to the same sets.

Examples of bounded ϕ^2 distributions are provided by the joint distribution of independent stochastic variables, in which case ϕ^2 is zero, and by the bivariate normal distribution with the absolute value of the correlation less than unity. All discrete distributions with finitely many points of increase in both variables will also have a finite ϕ^2. A case of special interest is provided by the bivariate joint normal distribution. In this distribution we may write $g(x)\,dx$ and $h(y)\,dy$ in place of $dG(x)$ and $dH(y)$ respectively and $f(x, y)\,dx\,dy$ in place of $dF(x, y)$. Pearson derived the relation,

$$(3) \qquad \phi^2 = \int\int f^2(x, y)/[g(x)h(y)]\,dx\,dy - 1 = \rho^2/(1 - \rho^2),$$

where $|\rho| < 1$. This result has been discussed by Lancaster [12]. However, if $|\rho| = 1$ and so the bivariate normal distribution is singular, ϕ^2 is unbounded. Indeed, ϕ^2 is unbounded for any bivariate distribution distributed along a straight line, with infinitely many points of increase.

It follows from the definition by an analysis similar to that used to justify the Riemann integral that ϕ^2 is uniquely determined by the passage to the limit if it is bounded.

DEFINITION. Let $\{x^{(i)}\}$ and $\{y^{(i)}\}$ be complete sets of orthonormal functions defined on the marginal distributions, $G(x)$ and $H(y)$, respectively by

$$(4) \qquad \int x^{(i)}x^{(j)}\,dG(x) = \int y^{(i)}y^{(j)}\,dH(y) = \delta_{ij}.$$

Let ρ_{ij} be the correlation coefficients,

$$(5) \qquad \rho_{ij} = \int\int x^{(i)}y^{(j)}\,dF(x, y).$$

By the Schwarz inequality ρ_{ij} always exists and is not greater than unity in absolute value. Further,

$$(6) \qquad \rho_{00} = 1, \qquad \rho_{0k} = \rho_{k0} = 0 \qquad k \neq 0.$$

The following discussion gives a statistical content to some well known analysis. The steps taken can be justified by the theory of integral equations as set out in Courant and Hilbert [2] or Riesz and Szent-Nagy [22].

THEOREM 1. *If $F(x, y)$ is a ϕ^2-bounded distribution and if*

$$(7) \qquad S_{mn} = S_{mn}(x, y) = \sum_{i=0}^{m}\sum_{j=0}^{n} \lambda_{ij}\, x^{(i)}y^{(j)},$$

then

$$(8) \qquad Q_{mn} = \int\int (\Omega - S_{mn})^2\,dG(x)\,dH(y)$$

is minimised by taking

$$(9) \qquad \lambda_{ij} = \rho_{ij}, \qquad i = 0, 1, 2, \cdots m; j = 0, 1, 2, \cdots n.$$

Writing S for S_{mn} as $m \to \infty$ and $n \to \infty$,

(10) $$\Omega(x, y) = S(x, y), \quad \textit{almost everywhere}$$

and

(11) $$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \rho_{ij}^2 = \phi^2.$$

PROOF. The set $\{x^{(i)}\} \times \{y^{(i)}\}$ is complete over the distribution $G(x) \times H(y)$, and $\Omega(x, y)$, as defined in (2), is square summable by (1B) and the hypothesis of the theorem. The result (9) follows by differentiating (7) with regard to λ_{ij} for $i = 0, 1, 2, \cdots m; j = 0, 1, 2, \cdots n$. For any finite m and n, the sum $\sum_{i,j} \rho_{ij}^2 \leqq \phi^2$, so that $\sum_{i,j} \rho_{ij}^2$ converges. The completeness assures the truth of (10) and of (11), which is the Parseval equality.

It is our aim now to redefine the sets $\{x^{(i)}\}$ and $\{y^{(j)}\}$ so that the correlation matrix,

(12) $$R = (\rho_{ij}), \qquad i = 1, 2, \cdots, j = 1, 2, \cdots,$$

assumes as simple a form as possible. The theorems of the next section show that R is diagonal if we choose, for the sets $\{x^{(i)}\}$ and $\{y^{(i)}\}$, the canonical variables in the sense of Fisher. The chief difficulty lies in the need to prove that the canonical variables form subsets of complete sets of orthonormal functions. We have, therefore, to proceed indirectly.

3. The Canonical Variables. The canonical variables have been defined on discrete distributions with finitely many points of increase. They are usually thought of as "scores to be assigned" but may also be thought of as functions of the marginal variables. Often no marginal variable has been explicitly defined; then, we may take the row or column position as the variable. The following definition may be regarded as the appropriate extension of Fisher's definition.

DEFINITION. The canonical variables (or functions) are two sets of orthonormal functions defined on the marginal distributions in a recursive manner such that the correlation between corresponding members of the two sets is maximal. Unity may be considered as a member of zero order of each set of variables. Symbolically, the orthogonal and normalising conditions are

(13) $$\begin{cases} \xi^{(i)} = \xi^{(i)}(x), \eta^{(i)} = \eta^{(i)}(y), & \\ \int \xi^{(i)}\, dG(x) = \int \eta^{(i)}\, dH(y) = 0, & i = 1, 2, \ldots, \\ \int \xi^{(i)2}\, dG(x) = \int \eta^{(i)2}\, dH(y) = 1, & i = 1, 2, \ldots, \\ \int \xi^{(i)}\xi^{(j)}\, dG(x) = \int \eta^{(i)}\eta^{(j)}\, dH(y) = 0 & \text{for } i \neq j, \end{cases}$$

and the maximisation conditions are that

$$\rho_i = \text{corr}\,(\xi^{(i)}, \eta^{(i)}) = \iint \xi^{(i)}\eta^{(i)}\, dF(x, y) \tag{14}$$

should be maximal for each i, given the preceding canonical variables. The ρ_i are the canonical correlations and can by convention be taken always to be positive.

THEOREM 2. *The canonical variables obey a second set of orthogonal conditions,*

$$E(\xi^{(i)}\eta^{(j)}) = \iint \xi^{(i)}\eta^{(j)}\, dF(x, y) = 0, \qquad \textit{if } i \neq j. \tag{15}$$

PROOF. For definiteness, let $j > i$. By hypothesis $E(\xi^{(i)}\eta^{(i)})$ is maximal in the sense of the definition above and is equal to ρ_i, say. Suppose that $E(\xi^{(i)}\eta^{(j)})$ is not zero but equal to $\rho_i \tan \theta$. Now $\eta^{(j)}$ has been defined according to (13) and so the function, $\text{Cos}\, \theta\eta^{(i)} + \text{Sin}\, \theta\eta^{(j)}$, obeys all the necessary orthogonal and normalising conditions, and its correlation with $\xi^{(i)}$ is easily found to be $\rho_i \sec \theta$ and this is greater than ρ_i, a contradiction results and so the theorem is proved.

As has been already noted, the canonical functions are necessarily square summable and so can be written as linear forms in any complete set of orthonormal functions, defined on the marginal distributions Thus we can write

$$\begin{cases} \xi^{(i)} = \sum_{k=1}^{\infty} a_{ik} x^{(k)}, & \sum_k a_{ik}^2 = 1, \\ \eta^{(i)} = \sum_{k=1}^{\infty} b_{ik} y^{(k)}, & \sum_k b_{ik}^2 = 1. \end{cases} \tag{16}$$

Now let us determine $\xi^{(1)}$ and $\eta^{(1)}$ in terms of the $\{x^{(i)}\}$ and $\{y^{(i)}\}$ respectively.

$$\begin{aligned} \text{Corr}\,(\xi^{(1)}, \eta^{(1)}) &= \text{corr}\,(\textstyle\sum_i a_i x^{(i)}, \sum_k b_k x^{(k)}) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_i b_j \rho_{ij}. \end{aligned} \tag{17}$$

Now $\sum_{i,j} \rho_{ij}^2$ is convergent and so the bilinear form on the right of (17) can be treated by the theory of quadratic forms in infinitely many variables. The normalising conditions (13) assure us that $\sum_i a_i^2 = 1$ and $\sum_j b_j^2 = 1$ and that neither $\xi^{(1)}$ nor $\eta^{(1)}$ contains any constant term. The bilinear form will have an attained maximum value for variations in the a_i and b_j. We take the coefficients of one such maximum to define a new set of variables

$$\begin{cases} x^{*(1)} = \xi^{(1)} = \sum_i a_i x^{(i)}, \\ x^{*(2)} = a_{21}x^{(1)} + a_{22}x^{(2)}, \\ x^{*(3)} = a_{31}x^{(1)} + a_{32}x^{(2)} + a_{33}x^{(3)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases} \tag{18}$$

where the a_{2j}, a_{3j}, $\cdots$ are chosen to satisfy the orthogonal and normalising conditions. A similar transformation is applied to the $y^{(i)}$:

$$(19)\qquad \begin{cases} y^{*(1)} = \eta^{(1)} = \sum_j b_j y^{(j)}, \\ y^{*(2)} = b_{21}y^{(1)} + b_{22}y^{(2)}, \\ y^{*(3)} = b_{31}y^{(1)} + b_{32}y^{(2)} + b_{33}y^{(3)}, \cdot \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

But now the correlation matrix, $R = (\rho_{ij})$, in the new variables is simpler in that, because of Theorem 2,

$$(20)\qquad \rho_{i1} = \rho_{1i} = 0 \qquad\qquad i \neq 1.$$

We can proceed similarly to find $\xi^{(2)}$ and $\eta^{(2)}$ in terms of the $\{x^{(i)}\}$ and $\{y^{(i)}\}$ respectively. Since $\xi^{(2)}$ is orthogonal to $\xi^{(1)}$

$$(21)\qquad \xi^{(2)} = \sum_2^\infty a_i^* x^{*(i)},$$

and similarly,

$$(22)\qquad \eta^{(2)} = \sum_2^\infty b_j^* y^{*(j)}$$

with $\sum_i a_i^{*2} = \sum_i b_i^{*2} = 1$. Now to find $\xi^{(2)}$ and $\eta^{(2)}$ we shall have to maximise $\sum_{i=2}^\infty \sum_{j=2}^\infty a_i^* b_j^* \rho_{ij}^*$. This again has an attained maximum and we take again a new set of variables

$$(23)\qquad \begin{aligned} x^{+(1)} &= x^{*(1)} = \xi^{(1)}, \\ x^{+(2)} &= \textstyle\sum_2^\infty a_i^* x^{*(1)} = \xi^{(2)}, \\ x^{+(3)} &= a_{32}^* x^{*(2)} + a_{33}^* x^{*(3)}, \\ x^{+(4)} &= a_{42}^* x^{*(2)} + a_{43}^* x^{*(3)} + a_{44}^* x^{*(4)}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots, \end{aligned}$$

and similarly define $y^{+(1)}$, $y^{+(2)}$, $y^{+(3)}$ $\cdots$ in terms of the $y^{(1)}$. The correlation matrix is simplified again for now

$$(24)\qquad \begin{cases} \rho^+_{1i} = \rho^+_{i1} = 0 & \text{for } i \neq 1, \\ \rho^+_{2i} = \rho^+_{i2} = 0 & \text{for } i \neq 2. \end{cases}$$

This process may be continued a denumerable infinity of times or until all ρ_i are zero for $i > r$ or $j > r$ for some value of r. We may follow Williams [25] and refer to r as the rank of the departure from independence. r may be infinite. At each step, since the transformation is orthogonal, a complete set is transformed into a complete set. It is evident that we may pass from the sets $\{x^{(i)}\}$ and $\{y^{(i)}\}$ by a series of orthogonal transformations to complete sets of orthonormal functions, of which the sets $\{\xi^{(i)}\}$ and $\{\eta^{(i)}\}$ are subsets and conversely. We can sum up these results in

THEOREM 3. *If $F(x, y)$ is a ϕ^2-bounded bivariate distribution with marginal distribution, $G(x)$ and $H(y)$, then complete sets of orthonormal functions can be defined on the marginal distributions such that each member of a set of canonical variables appears as a member of the complete set of orthonormal functions. The element of frequency can be expressed in terms of the marginal distributions,*

$$dF(x, y) = \left\{1 + \sum_1^\infty \rho_i x^{(i)} y^{(i)}\right\} dG(x)\, dH(y), \quad \text{a.e.,} \tag{25}$$

and

$$\phi^2 = \sum_{i=1}^\infty \rho_i^2 \tag{26}$$

PROOF. We have just proved the first statement. To prove the second we write, in the same way as in Theorem 1,

$$Q = \iint \{\Omega(x, y) - S_{mn}(x, y)\}^2\, dG(x)\, dH(y) \tag{27}$$

and take the partial differentials of Q with respect to λ_{ij}. Owing to the simplified form of the correlation matrix, ρ_{ij} is now zero for $i \neq j$ and ρ_{ii} is ρ_i. Since $\{x^{(i)}\} \times \{y^{(j)}\}$ is a complete set on $G(x) \times H(y)$, it follows that the minimised Q tends to zero as $m \to \infty$ and $n \to \infty$, and (26) which is the Parseval equality follows.

It may be proved that the choice of orthonormal functions is unique except for a convention as to sign if the ρ_i form a pair-wise different set. It is assumed throughout that, once $x^{(i)}$ is chosen, $y^{(i)}$ is defined so as to give the expectation of $x^{(i)}y^{(i)}$ a positive value. If, however, $\rho_{j+1}, \rho_{j+2}, \cdots \rho_{j+k}$ are of equal magnitude and $x^{(j+1)}, x^{(j+2)}, \cdots x^{(j+k)}$ is one solution for the corresponding canonical variables, then every other solution is given by an arbitrary orthogonal transformation on these $x^{(j+1)} \cdots x^{(j+k)}$ and the same transformation on the $y^{(j+1)} \cdots y^{(j+k)}$. A converse of Theorem 3 holds.

THEOREM 4. *If a bivariate distribution can be written in the form (25) with $\{x^{(i)}\}$ and $\{y^{(i)}\}$ forming complete sets on the marginal distribution and if $\sum_i \rho_i^2$ is finite, then the ρ_i are the canonical correlations, $x^{(i)}$ and $y^{(i)}$ are the canonical variables and $\sum_i \rho_i^2 = \phi^2$.*

PROOF. The proof is by induction. We suppose first that the ρ_i are pairwise different. Then if ξ and η are the first pair of canonical variables

$$\begin{aligned} \operatorname{corr}(\xi, \eta) &= \operatorname{corr}\left(\sum_i a_i x^{(i)}, \sum_j b_j y^{(j)}\right) \\ &= \sum_i a_i b_i \rho_i . \end{aligned} \tag{28}$$

Now $\sum_i a_i^2 = \sum_i b_i^2 = 1$ and Cauchy's inequality shows that the sum on the right of (28) is maximised by taking $a_1 = b_1 = 1$ and all other coefficients zero. Similarly, if $\rho_1 = \rho_2 \cdots = \rho_k$, Cauchy's inequality shows that the correlation of ξ and η is ρ_i if $\sum_1^k a_i^2 = 1$ and $a_i = b_i$ and that this is the maximum. Clearly however in this case too we can take $a_1 = b_1 = 1$, and once again $x^{(1)}$ and $y^{(1)}$ are the pair of first canonical variables or functions. We can proceed by induction

to prove the main statement of the theorem. Defining $\Omega(x, y)$ as in (2) and writing out its value by the use of (25), we derive

$$\sum_1^\infty \rho_i^2 = \phi^2.$$

This is a generalisation of a result of Hirschfeld [8] and Maung [13] in the finite case. Further, we may note that Theorem 3 is a generalisation of the Mehler identity; for, using the notation of (3), we define complete sets of orthogonal functions $\{x^{(i)}\} = \{\psi_i(x)\}$ and $\{y^{(i)}\} = \{\psi_i(y)\}$ on the marginal distributions where $\psi_i(x)$ is a polynomial of precise degree i standardised by the formula

(29)
$$\begin{cases} \int \psi_i(x)\psi_j(x)g(x)\, dx = \delta_{ij}, \\ \int \psi_i(y)\psi_j(y)h(y)\, dy = \delta_{ij}. \end{cases}$$

$g(x)$ and $h(y)$ have the same functional form in this case. By considering the expectation of $\exp\{tx - \frac{1}{2}t^2 + uy - \frac{1}{2}u^2\}$, namely $\exp \rho ut$, we find that

(30)
$$E\, x^{(i)}y^{(j)} = \delta_{ij}\rho^i$$

and Mehler's identity (Mehler [14]; Watson [24]) follows after Theorem 3 and continuity considerations. Conversely, given Mehler's identity, Theorem 4 shows that $|\rho|^i$ are the canonical correlations in this special case and the standardised Hermite-Chebishev polynomials, the canonical variables. Pearson [17] showed the great value of the Mehler identity in discussing normal correlation, although he and his collaborator, Bramley-Moore, failed to note that the tetrachoric expansion is indeed the Mehler identity. The Mehler identity is the special case when $f(x)$ and $g(y)$ are standardised normal distributions and $h(x, y)$ is the bivariate normal distribution with coefficient or correlation, ρ. This identity is given in Szegö's textbook [27] on page 371, where Szegö has $x\sqrt{2}$ and $y\sqrt{2}$ corresponding to our x and y and w for our ρ. Our $\psi_i(x)$ is $H_i(2^{-\frac{1}{2}}x)/\sqrt{i!}$ in his notation.

Dr. G. S. Watson (personal communication) has pointed out that the usual eigenfunction and kernel theory might be applied. The analogy is quite easy to establish in purely discrete or purely continuous distributions. In the continuous case we should define a kernel

(31)
$$K(x, y) = f(x, y)\, \{g(x)\, h(y)\}^{-\frac{1}{2}}$$

where $g(x) > 0$, $h(y) > 0$, with the convention that $K(x, y) = 0$ if $g(x)\, h(y) = 0$. $K(x, y)$ would in general be unsymmetric. It would follow that

(32)
$$\begin{cases} \rho_j y^{(j)}\sqrt{h(y)} = \int K(x, y)x^{(j)}\sqrt{g(x)}\, dx, \\ \rho_j x^{(j)}\sqrt{g(x)} = \int K(x, y)y^{(j)}\sqrt{h(y)}\, dy, \end{cases}$$

in precisely the same way as in equation (26) and (27) of Schmidt [23], noting the different definitions for the eigenvalues. (32) is proved by the application of Theorem 3. In the finite discrete case, where the frequencies are f_{ij}, the kernel $K(x, y)$ is replaced by $f_{ij}f_{i.}^{-\frac{1}{2}}f_{.j}^{-\frac{1}{2}} = b_{ij}$ and this is discussed in the next section. (32) is simplified if the marginal distributions are rectangular with $g(x) = h(y) = 1$

4. The Finite Case. The discussion above is a generalisation of a procedure, alternative to that of Fisher [3] and Maung [13], which may be used in the finite discrete case of an m by n contigency table with proportions f_{ij} in the cell of the ith row and jth column, with $f_{i.} = \sum_j f_{ij} > 0, f_{.j} = \sum_i f_{ij} > 0$, and for definiteness, $m \leq n$ It follows from Theorem 3 that if we construct matrices, X and Y, with the $(k + 1)$th column consisting of the values of the kth canonical variable, then $X'FY$ will have a canonical form with non-zero elements everywhere except along the leading diagonal. It is found simpler to deal with a matrix B derived from F and then the problem is reduced to determining a canonical form for a rectangular matrix under pre- and post-multiplication by orthogonal matrices, which we consider by an adaptation of the argument of Murnaghan [15] on his pages, 26 and 27. The defining conditions for the matrices X and Y may be written

$$(33)\qquad \begin{aligned} x_{i1} &= 1 & i &= 1, 2, \cdots m, \\ y_{i1} &= 1 & i &= 1, 2, \cdots n, \\ x_{ij} &= \xi_{(i)}^{(j-1)} = \xi_i^{(j-1)}, & j &= 2, 3, \cdots, m, \\ y_{ij} &= \eta_{(i)}^{(j-1)} = \eta_i^{(j-1)}, & j &= 2, 3, \cdots, n \end{aligned}$$

(13) now becomes

$$(34)\qquad \begin{cases} X' \operatorname{diag} f_{i.} X = 1_m , \\ Y' \operatorname{diag} f_{.j} Y = 1_n , \end{cases}$$

and the elements of the leading diagonal of $X'FY$ are to be maximised. Theorem 4 ensures that it is sufficient and Theorem 2 that it is necessary for $X'FY$ to be in canonical form. We therefore state without completing the proof

THEOREM 5. *Given an $m \times n$ contigency table with proportions f_{ij} in the cell of the* ith *and* jth *column, let an $m \times n$ matrix, B, be defined by*

$$(35)\qquad b_{ij} = f_{ij}f_{i.}^{-\frac{1}{2}}f_{.j}^{-\frac{1}{2}}$$

Then orthogonal matrices M and N exist with elements of the first column $\sqrt{f_{i.}}$ and $\sqrt{f_{.j}}$, respectively such that $M'BN$ is in canonical form, namely

$$(36)\qquad M'BN = C = [\operatorname{diag}(1, \rho_1 \cdots \rho_{m-1}), 0_{m,n-m}].$$

It is evident further by a consideration of the forms of $(M'BN)$ $(M'BN)'$ and $(M'BN)'$ $(M'BN)$ that M and N are the orthogonal matrices that reduce BB' and $B'B$ respectively to canonical form. Conversely, it can be shown that if N

transforms $B'B$ to canonical form with unity in the leading position and k other non-zero diagonal elements, then an M, having for its first $(k+1)$ columns the first $(k+1)$ columns of BN normalised, can be constructed so that $M'BN$ is in the required form. In fact, the first $(k+1)$ columns of BN are mutually orthogonal because $(NB)'(BN)$ is diagonal. Maung [13], obtains the latent roots of BB' or $B'B$ by solving the determinantal equation, $| BB' - \lambda 1 | = 0$, in the usual manner. An alternative is to use the iteration method of Frazer, Duncan and Collar ([6], page 133). We note further that M and N must be of the form,

$$\text{(37)} \qquad \begin{cases} M = M_1(1 \dotplus M_2), \\ N = N_1(1 \dotplus N_2), \end{cases}$$

where M_1 and N_1 are of the Helmert type with first columns having elements $f_{i.}^{\frac{1}{2}}$ and $f_{.j}^{\frac{1}{2}}$ respectively. Now the elements of

$$\text{(38)} \qquad M_1'BB'M_1 = 1 \dotplus W$$

can be computed readily. Using the observed number, a_{ij} in the contigency table,

$$\text{(39)} \qquad W_{kk'} = \frac{a_{..}\left(a_{k+1,.}\sum_{i=1}^{k} a_{ij} - a_{k+1,j}\sum_{i=1}^{k} a_{i.}\right)\left(a_{k'+1,.}\sum_{i=1}^{k'} a_{ij} - a_{k'+1,j}\sum_{i=1}^{k'} a_{i.}\right)}{a_{.j}\left\{a_{k.}\,a_{k'}\,a_{k+1,.}\,a_{k'+1,.}\sum_{1}^{k} a_{i.}\sum_{1}^{k'} a_{i.}\right\}^{\frac{1}{2}}}$$

The trace of W is χ^2. It does not take much more time to compute W than χ^2 if m is not too large. A computing routine is to form a matrix with elements in the first row, $(a_{1j}a_{2.} - a_{1.}a_{2j})$, elements in the second row $(a_{1j} + a_{2j})a_{3.} - (a_{1.} + a_{2.})a_{3j}$ and so on. For each row, a standardising factor is computed,

$$\left\{a_{k.}\,a_{k+1,.}\sum_{1}^{k} a_{i.}\right\}^{\frac{1}{2}}.$$

The elements of W are then simply computed by formula (39). The Helmert matrix can be looked upon as generating sets of orthonormal functions, which take a simple form. The values for the canonical variables are then calculated by an orthogonal transformation

$$\text{(40)} \qquad \begin{aligned} X &= \operatorname{diag} f_{i.}^{-\frac{1}{2}} M \\ &= f_{i.}^{-\frac{1}{2}} M_1(1 \dotplus M_2) \end{aligned}$$

where M_1 is the Helmert Matrix and $M_2'WM_2$ is diagonal, M_2 being obtained by iteration and similarly Y can be written in terms of N_1 and N_2.

A Numerical Example. Maung [13] has given the following example of a classification of Aberdeen schoolchildren by hair and eye colours (see Table I).

A matrix of elements, U, with $u_{kj} = (a_{k+1,.}\sum_{i=1}^{k} a_{ij} - a_{k+1,j}\sum_{i=1}^{k} a_{i.})$ is given by

$$\begin{bmatrix} 1{,}487{,}190 & -273{,}082 & -1{,}077{,}957 & -110{,}090 & -26{,}061 \\ 16{,}182{,}645 & 773{,}584 & -8{,}895{,}366 & -7{,}831{,}720 & -229{,}143 \\ 19{,}806{,}181 & 1{,}123{,}770 & 7{,}415{,}022 & -26{,}653{,}016 & -1{,}691{,}957 \end{bmatrix}.$$

TABLE I

Eye colour	Hair colour					Total
	Fair	Red	Medium	Dark	Black	
Blue	1368	170	1041	398	1	2978
Light	2577	474	2703	932	11	6697
Medium	1390	420	3826	1842	33	7511
Dark	454	255	1848	2506	112	5175
Total	5789	1319	9418	5678	157	22,361

The elements of this matrix are now divided by the corresponding column totals of the contigency table to give a matrix (v_{ij}). Divisors appropriate to each row of U are now computed, $\{a_{k.}a_{k+1.}\ \sum_1^k a_{i.}\}^{\frac{1}{2}} = d_k$ Then w_{ij} is $\sum_k u_{ik}v_{jk}/\{d_i d_j\}$ or $\sum_k v_{ik}u_{jk}/\{d_i d_j\}$. We thus obtain the matrix, W, of (38).

$$\begin{bmatrix} 65.8744811 & 237.1027158 & 173.4280109 \\ 237.1027158 & 1167.9147643 & 1252.2082711 \\ 173.4280109 & 1252.2082711 & 2450.0865906 \end{bmatrix}.$$

The trace of W is 3683.875836 agreeing with Maung's value for χ^2.

The orthogonal matrix, M_2, of (37) is then derived from W by an iteration process and is

$$\begin{bmatrix} 0.085413 & 0.272546 & 0.958344 \\ 0.522636 & 0.806650 & -0.275985 \\ 0.848266 & -0.524438 & 0.073545 \end{bmatrix}.$$

The values of the complete set of orthonormal variables associated with the Helmert matrix, M_1, may be displayed as a matrix,

$$\begin{bmatrix} 1 & 2.279806 & 1.005036 & 0.548741 \\ 1 & -1.013777 & 1.005036 & 0.548741 \\ 1 & 0 & -1.294598 & 0.548741 \\ 1 & 0 & 0 & -1.822352 \end{bmatrix}.$$

In the jth column, all elements above the diagonal are equal to $\{p_{j.}/(\sum_1^{j-1} p_{k.}\sum_1^j p_{k.})\}^{\frac{1}{2}}$, the diagonal element is $-\{\sum_1^{j-1} p_{k.}/(p_{j.}\sum_1^j p_{k.}\}^{\frac{1}{2}}$ and element below the diagonal are zero. Post-multiplication of this matrix by $(1 \dotplus M_2)$ yields the sets of canonical variables in the form of a 4×4 matrix, X, of Equation (40)

$$\begin{bmatrix} 1 & +1.1855 & +1.1443 & +1.9478 \\ 1 & +0.9042 & +0.2466 & -1.2086 \\ 1 & -0.2111 & -1.3321 & +0.3976 \\ 1 & -1.5458 & +0.9557 & -0.1340 \end{bmatrix}.$$

The values of the elements agree with those given by Maung.

The canonical variables in y can now be obtained by using Fisher's algorithm

as in (45), below, and we may write the first four columns of the matrix, Y, as

$$\begin{bmatrix} 1 & +1.3419 & +0.9713 & +0.3288 \\ 1 & +0.2933 & -0.0236 & -3.7389 \\ 1 & +0.0038 & -1.1224 & +0.1666 \\ 1 & -1.3643 & +0.7922 & +0.3625 \\ 1 & -2.8278 & +3.0607 & -3.8177 \end{bmatrix}.$$

Programs, similar to the computational process used above, are now available on electronic computers.

Interpreting the findings, the first set of canonical variables arranges both hair colour and eye colour in the same order as was suggested by biological considerations. If there is an underlying bivariate distribution the first set of canonical variables gives the best values to be assigned to the marginal variables.

5. Identifications of the Finite and the General Cases. We now state some corollaries deducible from the theorems above in such a way as to bring out the identity of the theory of canonical correlation as a special case of the more general theory; where appropriate, we have numbered these "a" for the finite case, "b" for the more general.

COROLLARIES.

(ia). ρ_i^2 are the non-zero latent roots of the matrices BB' and BB'; ρ_i are the "roots" of B under transformation by pre- and post-multiplication of B by orthogonal matrices.

(ib). ρ_i^2 are the eigenvalues of certain symmetric kernels and ρ_i are the eigenvalues of a certain, possibly asymmetric, kernel.

(iia). The identity of Fisher [4]

$$f_{ij} = f_{i.}f_{.j}\left\{1 + \sum_{k=1}^{m-1} \rho_k x^{(k)} y^{(k)}\right\} \tag{41}$$

is a special case of our Theorem 3. It is also proved by noting that

$$X'AY = M'BN = C, \tag{42}$$

and the inverse of X' is diag $f_{i.}X$ and the inverse of Y is Y' diag $f_{.j}$ by (34).

(iib). The generalisation of Fisher's identity is given by Theorem 3.

(iiia) and (iiib). If $\mathbf{m}_k$ and $\mathbf{n}_k$ are the kth column vectors of M and N respectively

$$\begin{cases} \rho_k \mathbf{n}_k = B'\mathbf{m}_k\,, \\ \rho_k \mathbf{m}_k = B\mathbf{n}_k\,, \end{cases} \tag{43}$$

or alternatively after (36)

$$\begin{cases} BN = MC, \\ B'M = NC', \end{cases} \tag{44}$$

or

$$\begin{cases} AY = \operatorname{diag} f_{i.}XC, \\ A'X = \operatorname{diag} f_{.j}YC', \end{cases} \tag{45}$$

(45) corresponds exactly with equations (26) and (27) of Schmidt [23] as modified in our (32). The equation (45) is the basis of Fisher's [3] algorithm for the computation of the canonical correlations, which we give as a corollary.

(iv). The canonical variables can be obtained by iteration if $\rho_{j+1} > \rho_{j+2}$. From (45) it follows that

$$\text{(46)} \qquad \operatorname{diag} f_{i.}^{-1} A \operatorname{diag} f_{.i}^{-1} A' X = X C C',$$

and so

$$\text{(47)} \qquad (\operatorname{diag} f_{i.}^{-1} A \operatorname{diag} f_{.i}^{-1} A')^p X \mathbf{n}_0 = X (C C')^p \mathbf{x}_0 .$$

Therefore if any vector $\mathbf{x}_0$ is taken orthogonal to the first j columns of X but not orthogonal to the $(j + 1)$th column, the iteration of the form (45) will yield a vector proportional to the $(j + 1)$th column of X. This is a special case of iterating using Schmidt's (26) and (27), which we could rewrite as (ivb).

(v). In Yates [26], arises the problem to find values for y such that y will have maximum correlation with an x, which has prescribed values

We may write

$$\text{(48)} \qquad x = \sum_{i=1}^{m-1} a_i x^{(i)}, \qquad \sum_i a_i^2 = 1.$$

Then from the canonical form of Theorem 3 and the use of the Cauchy inequality, we find that

$$\text{(49)} \qquad y = \sum_{i=1}^{m-1} a_i \rho_i y^{(i)},$$

is such that the correlation of x and y is maximal and

$$\text{(50)} \qquad \operatorname{corr}(x, y) = \left(\sum_{i=1}^{m-1} a_i^2 \rho_i^2 \right)^{\frac{1}{2}}.$$

(vi). In either finite or infinite cases, it can be proved that the existence of k canonical correlations of unity means the distribution consists of $(k+1)$ disjunct pieces. The case of one canonical correlation of unity has been treated by Richter [21].

6. Regression in the Bivariate Distribution. If the bivariate surface can be described in the canonical form (25), then regression takes a particularly simple form.

THEOREM 6. *The regressions of the canonical variables are given by the lines,*

$$\text{(51)} \qquad \begin{aligned} x^{(i)} &= \rho_i y^{(i)}, \\ y^{(i)} &= \rho_i x^{(i)}. \end{aligned}$$

For $i \neq j$ the regression of $x^{(i)}$ on $y^{(j)}$ and $y^{(i)}$ on $x^{(j)}$ are zero.

PROOF. This follows in the usual way by minimising

$$\iint (x^{(i)} - \lambda y^{(j)})^2 \, dF(x, y).$$

identally, we have proved that the regression of $x^{(i)}$ on $y^{(i)}$ is linear since any are summable function of $y^{(i)}$ orthogonal to $y^{(i)}$ can be expanded in terms of other orthonormal functions.

. Generalization of the Notion of Correlation. Many attempts have been de to find some way of obtaining bivariate distributions which would generalthe normal case. Pretorius [20] has given many references to such attempts. her's theory of canonical correlation gives an alternative approach. Suppose are given marginal variables with distribution functions $G(x)$ and $H(y)$, then bivariate distribution can be formed using (25) provided that the series $+ \sum_1^\infty \rho_i x^{(i)} y^{(i)}]$ is non-negative at points corresponding to increase in both) and $H(y)$. We may take one of the simplest possible pairs of distributions the margins, namely the rectangular over the range $-\frac{1}{2}$ to $\frac{1}{2}$ and set up three erent bivariate distributions.

EXAMPLE 1. We take as our orthonormal sets of functions the normalised gendre polynomials, in particular

$$\begin{cases} x^{(1)} = x\sqrt{12}, \\ x^{(2)} = 6\sqrt{5}\,(x^2 - \frac{1}{12}). \end{cases}$$

can now assign correlations ρ_1 and ρ_2 subject to the condition that the density omes nowhere negative

$$dF(x, y) = \{1 + 12\rho_1 xy + 180\rho_2(x^2 - \tfrac{1}{12})(y^2 - \tfrac{1}{12})\}\, dx\, dy.$$

the maximum absolute value of $x^{(1)}y^{(1)}$ is 3 and that of $x^{(2)}y^{(2)}$ is 5, so the ression in (53) will be positive if

$$3\,|\,\rho_1\,| + 5\,|\,\rho_2\,| < 1.$$

EXAMPLE 2. We choose the cosine series as the orthonormal sets,

$$\begin{cases} x^{(1)} = \sqrt{2}\,\mathrm{Cos}(2\pi x), \\ x^{(2)} = \sqrt{2}\,\mathrm{Cos}(4\pi x), \end{cases}$$

l similarly define $y^{(1)}$ and $y^{(2)}$

$$dF(x, y) = \{1 + 2\rho_1\,\mathrm{Cos}\,(2\pi x)\,\mathrm{Cos}\,(2\pi y) + 2\rho_2\,\mathrm{Cos}\,(4\pi x)\,\mathrm{Cos}\,(4\pi y)\}\, dx\, dY.$$

is is non-negative if the absolute value of ρ_1 and ρ_2 are both less than $\frac{1}{2}$.

EXAMPLE 3. A further possibility results from forming arbitrary bivariate tributions, e.g., we might divide the square with corners at $(+\frac{1}{2}, +\frac{1}{2})$ into r quarters and add $+\rho_1$ to the density in the first and third quadrants and otract ρ_1 from the density in the second and fourth quadrants. We could also odivide the original square into 16 parts and add ρ_2 to the four corner subrisions and to the four central subdivisions and subtract ρ_2 from the remainder.

The resulting distribution can be described with the aid of step-functions

$$dF(x, y) = \{1 + \rho_1 x^{(1)} y^{(1)} + \rho_2 x^{(2)} y^{(2)}\}\, dx\, dy, \tag{57}$$

where

$$\begin{cases} x^{(1)} = +1 \text{ for } x \leqq 0, \text{ for } x < 0, \\ x^{(2)} = -1 \text{ for } \frac{1}{4} \leqq x < \frac{1}{4} \text{ and } +1 \text{ for } x \text{ elsewhere.} \end{cases} \tag{58}$$

To obtain a complete set of orthogonal functions defined on $[-\frac{1}{2}, \frac{1}{2}]$ we divide this interval into four subintervals of equal length. On each complete sets of orthonormal functions may be defined. For example, we may choose the Legendre polynomials as our set, standardized so as to be orthonormal on the uniform distribution $[-\frac{1}{2}, \frac{1}{2}]$. Corresponding to the first interval we define a set of orthogonal polynomials which have the values $1 = P^{(0)}(X)$, $P^{(i)}X$, $i = 1, 2 \cdots$ where $X + \frac{1}{2}$ is the fractional part of $4(x + 1)$, on the first interval and zero elsewhere and similar sets on the other subintervals. The four sets of functions may be displayed as the elements of a four rowed matrix, P, of infinitely many columns The rows of this matrix are obviously mutually orthogonal since no two elements of the same column can be simultaneously non-zero Let us now define $Q = AP$, where A is the matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$$

The elements of Q are now an orthonormal set on the whole interval $q_{11} = x^{(0)}$, a term constant on $[\frac{1}{2}, \frac{1}{2}]$. $q_{21} = x^{(1)}$, $q_{31} = x^{(2)}$. q_{41} is necessary for completeness. It is constant on any subinterval but changes sign being -1 on the odd intervals Every other function q_{ij} of the form $\pm P^{(i)}(X)$. The $y^{(i)}$ may be similarly defined

It is clear from the examples that the same correlations can arise in a great many different ways. In the next section, we show how the methods can be used as a test of normality.

These three examples show how bivariate distributions can be formed with arbitrarily prescribed correlation coefficients. Barrett and Lampard [1] give two other examples where such bivariate distributions arise naturally out of a physical problem.

8. A Canonical Partition of χ^2. In testing whether a bivariate distribution is normal, the marginal distributions can be tested in the usual way by an overall χ^2 or the individual degrees of freedom can be displayed as previously suggested by Lancaster [11] by the aid of orthogonal polynomials. Moreover, according to the analysis of the present paper and that of Lancaster [11] the regressions of the orthogonal polynomials in x and y on one another should be zero except for polynomials of the same degree We therefore may compute the regres-

sions and display them in the form of a matrix, which we explain with the aid of a well-known example, the correlation table of Pearson and Lee (*Biometrika* 2,257), easily accessible in ([5], paragraph 30). After estimating the mean and variance of both variables, the regressions of the theoretical Hermite-Chebishev in one variable on those of the other may be computed and set out as suggested by Lancaster [11]. The mean and standard deviations have been computed using n as a divisor. The table of Pearson and Lee has been modified to 8 columns representing classifications of daughters' heights. The $\psi_i(x)\psi_j(y)$ sums of products of polynomials of the form, $\psi_i(x)\psi_j(y)f_{ij}$, have been computed and divided by 1376 the number of observations to give component χ's of a partition of χ^2. The leading 4×4 submatrix is as follows—

$$\begin{bmatrix} \cdot & \cdot & \cdot & -1.006 \\ \cdot & 19.238 & -0.053 & -1.834 \\ \cdot & 0.398 & 8.325 & -0.460 \\ -0.328 & -0.578 & -0.350 & 2.390 \end{bmatrix}$$

The term 19.238 corresponds to the regression of the first polynomial in the fathers' heights on first polynomial in the daughters' heights and to a correlation of 0.5186, which is slightly different from that given by Fisher [5] as the grouping is different. It may be noted also that the squares of the 3×3 submatrix excluding the marginal terms accounts for over 446 of a χ^2 of 504.23 if the table is analysed by the usual χ^2 with fixed marginal totals, so that all the significant departure from independence is shown to be accounted for by the first three not identically zero diagonal terms, the sum of whose squares is 445.

Pearson [19] gave a rule which substantially states that the number of degrees of freedom must be subtracted from the χ^2 of the test of homogeneity when computing ϕ^2. We have

$$\begin{aligned} \phi^2 &= (504.234 - 98)/1376 \\ &= 0.295228, \\ \rho^2 &= 0.295228/1.295228 \\ &= 0.227935, \\ \rho &= 0.477, \end{aligned}$$

which gives a correlation approximately equal to that calculated here, 0.5186.

An alternative canonical partition is given by estimating the means and variances and computing the marginal frequencies on the assumption of normality. A partition of χ^2 is obtained as shown in Table II.

It is clear that the distribution of Pearson and Lee is fitted very well by the assumption that it is a sample of a bivariate normal distribution. The residual χ^2 of 101.04 with 95 degrees of freedom represents the sums of squares due to all other regressions than the first three regressions of the form $\psi_i(x)$ on $\psi_i(y)$. The assumption of normality of the marginal distributions and a non-zero correlation are sufficient to account for the total χ^2, for the residual χ^2 is little greater than the corresponding degrees of freedom.

TABLE II

Source of χ^2	Degrees of Freedom	χ^2
Difference of distribution of father's heights from theoretical	5	7.20
Difference of distribution of daughter's heights from theoretical	12	12.77
Regression of $\psi_1(y)$ on $\psi_1(x)$ . .	1	370.10
Regression of $\psi_2(y)$ on $\psi_2(x)$.	1	69.31
Regression of $\psi_3(y)$ on $\psi_3(x)$.	1	5.71
Residual . .	95	101 04
Total	115	566 13

9. Summary. The problems of Hirschfeld [8] and of the description of a contingency table by means of the canonical variables and correlations have been generalised to distributions limited only by the condition that the Pearson ϕ^2 is finite. Any theoretical or observed distribution subject to this condition can be described by the canonical variables (that is, subsets of complete sets of orthogonal functions in the variables of the two marginal distributions, which obey the second orthogonality condition that $Ex^{(i)}y^{(j)}$ is zero for $i \neq j$, and the canonical correlations. The theory of Fisher [3], Maung [13] and Williams [25] has been related to the eigenfunction theory.

Mehler's identity, or in statistical language, the expansion of the bivariate normal frequency in tetrachoric functions, has been generalised. The approach of Maung [13] has been modified to allow for an extension of the canonical theory to continuous marginal distributions.

The methods used give a new test of goodness of fit for the bivariate normal distribution and enable populations to be constructed with arbitrary marginal distributions and correlations.

Acknowledgments. This paper is published with the permission of the Director-General of Health, Canberra.

Dr. G. S. Watson has offered valuable help in its preparation.

REFERENCES

[1] J. F. BARRETT, AND D. G LAMPARD, "An expansion for some second-order probability distributions and its application to noise problems," IRE *Transactions of the Professional Group on Information Theory*, IT 1 (1955), pp. 10–15.

[2] R. COURANT, AND D. HILBERT, *Methods of Mathematical Physics*, Interscience Publishers, Inc., New York, 1953.

[3] R. A. FISHER, "The precision of discriminant functions," *Ann. Eugenics*, Vol 10, (1940), pp. 422–429.

[4] R. A. FISHER, cited by Maung [13].

[5] R. A. FISHER, *Statistical Methods for Research Workers*, 10th ed , Oliver and Boyd, Edinburgh, 1944.

[6] R. A. FRAZER, W. J. DUNCAN, AND A. R. COLLAR, *Elementary Matrices*, Cambridge University Press, 1946.

[7] E. HELLINGER, "Neue Begruendung der Theorie quadratischer Formen von unendlichvielen Veraenderlichen", *J. Reine Angew. Math.* (1909), Vol. 136, pp. 210-271.
[8] H. O. HIRSCHFELD, "A Connection between Correlation and Contingency", *Proc. Cambridge Philos. Soc.* Vol. 31 (1935), 520-524.
[9] E. W. HOBSON, "On Hellinger's Integrals", *Proc. London Math. Soc.* (2), Vol. 18 (1919), pp. 249-265.
[10] H. HOTELLING, "Relations between two Sets of Variates", *Biometrika*, Vol. 28 (1936), pp. 321-377.
[11] H. O. LANCASTER, "A Reconciliation of χ^2 Considered from Metrical and Enumerative Aspects", *Sankhyā*, Vol. 13 (1953), pp. 1-10.
[12] H. O. LANCASTER, "Some properties of the bivariate normal distribution considered in the form of a contingency table", *Biometrika*, Vol. 44 (1957), pp. 289-292.
[13] K. MAUNG, "Measurement of Association in a Contingency Table with Special Reference to the Pigmentation of Hair and Eye Colours of Scottish School Children", *Ann. Eugenics*, Vol. 11 (1941), pp. 189-223.
[14] F. G. MEHLER, "Ueber die Entwicklung einer Funktion von beliebig vielen Variablen nach Laplaceschen Funktionen hoeherer Ordnung", *J. Reine Angew. Math.*, Vol. 66 (1866), pp. 161-176.
[15] F. D. MURNAGHAN, *The Theory of Group Representations*, The Johns Hopkins Press, Baltimore, 1938.
[16] K. PEARSON, "On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling", *Philos. Mag.* (5), Vol. 50 (1900), pp. 157-175.
[17] K. PEARSON, "Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable", *Philos. Trans. Roy. Soc. (Lond.)* Vol. 195A (1900), pp. 1-47.
[18] K. PEARSON, "Mathematical contributions to the theory of evolution. XIII. On the theory of contingency and its relation to association and normal correlation", *Drapers Company Research Memoirs Biometric Series, No. 1* (1904)
[19] K. PEARSON, "On the Measurement of the Influence of Broad Categories on Correlation", *Biometrika*, Vol. 9 (1913), pp. 116-139.
[20] S. J. PRETORIUS, "Skew bivariate frequency surfaces examined in the light of numerical illustrations", *Biometrika*, Vol. 22 (1930), pp. 109-223.
[21] H. RICHTER, "Zur Maximalkorrelation", *Z. angew. Math. Mech.* Vol. 29 (1949), pp. 127-128.
[22] F. RIESZ, AND B. SZENT-NAGY, *Functional Analysis*. Translated from the 2d French Edition, 1956.
[23] E. SCHMIDT, "Zur Theorie der linearen und nichtlinearen Integralgleichungen", *Math. Ann.*, Vol. 63 (1907), pp. 433-476.
[24] G. N. WATSON, "Notes on generating functions of polynomials: Hermite polynomials", *J. London Math. Soc.*, Vol. 8 (1933), pp. 194-199.
[25] E. J. WILLIAMS, "Use of scores for the analysis of association in contingency tables", *Biometrika*, Vol. 39 (1952) pp. 274-289.
[26] F. YATES, "The analysis of contingency tables with groupings based on quantitative characters", *Biometrika*, Vol. 35 (1948), pp. 176-181.
[27] G. SZEGÖ, (1939), *Orthogonal Polynomials*, Colloquium Publ. No. 23, Amer. Math. Soc.

ON RENEWAL PROCESSES RELATED TO TYPE I AND TYPE II COUNTER MODELS[1]

By Ronald Pyke

Stanford University

Summary. Several renewal processes related to the Type I and Type II counter models are defined and studied. The distribution and characteristic functions for the secondary (or output) process of the Type I counter model are obtained explicitly. Both the non-stationary and stationary probabilities of the state of the counter, (locked or unlocked), are derived. Integral equations determining the distribution and characteristic functions for the secondary process of the Type II counter model are obtained. Also it is shown that a more general model proposed by Albert and Nelson [1] may be solved explicitly in terms of a corresponding Type II counter model. An example of this general model is given. Related with each model is a discrete renewal process which is also studied.

1. Introduction and Notation. Two important classes of counting devices are the Type I and Type II counters defined as follows. A counter for detecting radioactive impulses is placed within range of a radioactive material. By "an event has happened", we mean that an impulse has been emitted by the material and by "an event has been registered", we mean that an impulse emitted by the material has been detected and recorded by the counter. Due to the inertia of the counting device, all impulses will probably not be counted. The time during which the device is unable to record an impulse is referred to as deadtime.

Definition. A Type I counter is one in which deadtime is produced only after an event has been registered. A Type II counter is one in which dead time is produced after each event has happened. Examples of Type I and Type II counters are the Geiger-Müller counters and electron multipliers respectively.

In sections 4 to 7, attention will be given only to the Type I problem. It is stated theoretically as follows. Let X, Y and Z be random variables (r.v.) with distribution functions (d.f.) F, G and H respectively. Let $\{X_i\}_{i=1}^{\infty}$, $\{Y_j\}_{j=0}^{\infty}$ be independent X- and Y-renewal processes; that is $\{X_i, Y_j: i \geqq 1, j \geqq 0\}$ is a family of mutually independent r.v.'s and each X_i and Y_j has d.f. F and G respectively. Set $X_0 = 0$ (a.s.) and $S_k = \sum_{i=0}^{k} X_i$ for $k = 0, 1, 2, \cdots$. Assume throughout this discussion that $F(0) = G(0-) = 0$, F is a non-lattice distribution and that all d.f.'s are right continuous. Define $n_0 = 0$ and

$$n_j = \min\{k \in I^+ : S_k > Y_{j-1} + S_{n_{j-1}}\}$$

for $j = 1, 2, 3, \cdots$, where I^+ is the set of positive integers. The above definitions are valid with probability one.

Received October 29, 1957; revised March 28, 1958.

[1] This work was sponsored in part by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

The secondary renewal process, $\{Z_i\}_{i=1}^{\infty}$ (to be referred to as the Z-process) is defined by

$$Z_i = S_{n_i} - S_{n_{i-1}} \qquad (i \varepsilon I^+).$$

This is clearly a renewal process since the S_j's are sums of independent r.v.'s and since $\{n_j - n_{j-1}\}_{j=1}^{\infty}$, a sequence of identically and independently distributed r.v.'s, is itself a renewal process. $\{n_j - n_{j-1}\}_{j=1}^{\infty}$ shall be referred to as the N-process, and H shall denote the common c.d.f. of the Z-process. It will be shown that $E(n_1)$ denotes the asymptotic bias of the counter.

One may define a related stochastic process which is of interest in counting problems. Let $\{V_t : t \geqq 0\}$ be a stochastic process, having a two point range space, with joint distribution functions derived from its definition which is: $V_0 = 0$ (a.s.) and

$$V_t = \begin{cases} 1 \text{ if } Z_k + Y_k \leqq t < Z_{k+1} \text{ for some } k \varepsilon I^+ \\ 0 \text{ otherwise} \end{cases}$$

Set

$$P_1(t) = 1 - P_0(t) = \Pr[V_t = 1]$$

and

$$P_1 = 1 - P_0 = \lim_{t \to \infty} P_1(t)$$

if the limit exists.

A subscript, j say, affixed to any distribution function will denote its jth convolution with itself. The zero subscript will denote the c.d.f. degenerate at zero.

In sections 8 and 9 the Type II problem is studied. Its theoretical formulation differs from the Type I problem only in the definition of the N-process, which for the Type II problem is $n_0 = 0$ and

(1) $$n_j = \min\{k \varepsilon I^+ : k > n_{j-1}, S_k > S_r + Y_r, r = n_{j-1}, \cdots, k-1\}.$$

In all other instances, the definitions remain unchanged. For example, the secondary renewal process is still given by

$$Z_i = S_{n_i} - S_{n_{i-1}} \qquad (i \varepsilon I^+),$$

although, it is clearly a different process. The same notation is used for both models in order to emphasize to the reader the common interpretation of the various symbols.

In section 10 a more general model, suggested by Albert and Nelson [1], is studied. It is shown that the solution of this more general model is an immediate consequence of the solution of a corresponding Type II problem.

We shall begin in section 3 by proving a theorem from which the quantities $P_1(t)$ are immediately deducible.

To understand the connection between the above notation and the counter

problem itself, let Y_j represent the deadtime caused respectively by the registration of an event at time S_{n_j} in the Type I model and the happening of an event at time S_j in the Type II model (time being measured from the registration of some event) and let X_k be the time between the kth and $(k + 1)$-st impulses. The secondary renewal process is determined by the r.v. Z, which denotes the time between successive counts or registrations. The event $[V_t = 1]$ corresponds to the counter being unlocked at time t. For a more detailed description of the physical problem, the reader is referred to the references. (See e.g., Feller [2].)

Throughout this paper, the integrals that appear are to be considered as Lebesque-Stieltjes integrals. This will avoid the special considerations that would otherwise be required in cases where the integrand has a set of discontinuities of positive measure with respect to the Stieltjes measure. Notice that the ordinary integration-by-parts formula holds for the Lesbesque-Stieltjes integrals that appear in this paper. A proof of this is possible by probabilistic methods.

2. The literature and known results. The Type I and Type II counter problems have been studied by several people. Most of these studies deal with the special case in which the input process is Poisson. Not only does the Poissonian input make the problems involved more tractable, but in this instance, it serves to make the statistical model very realistic, since the impulses from a radioactive material behave randomly over time, at least in time intervals which are short relative to the half-life of the material. For an extensive bibliography, the reader is referred to Takacs [3].

It is important, however, to study the more general non-poissonian models for several reasons. First of all, it is necessary at times to make successive counts and it is known that the secondary process of the first counter, which would serve as the input process for the second counter, is not a Poisson process even though the original process was. Secondly, these same theoretical models have arisen in other contexts in which the Poisson process is not so easily justified (e.g., in inventory theory, Arrow, Karlin and Scarf [4]).

In his recent paper, [5], received by this author after completion of the first draft of this paper, Takacs also studies the general counter problem. Although there is some overlap, there are many differences in approach and coverage between the two treatments of the problem. Theorem 2 is equivalent to results obtained by Tackacs in [3] and again in [5], for the case of continuous F and G. Even for this case, however, our result (4) is a simplification in that a double integral has been replaced by a single one. Attention should also be given to a recent paper of Smith [6], in which the Type II counter model with Poissonian input (and related quasi-Poissonian inputs) as well as the model with constant deadtime, is studied.

3. A related renewal problem. In this section, we shall consider two alternating renewal processes, not necessarily independent, and obtain explicitly the probabilities, both finite and stationary, of one of the processes being in effect at any

given instant of time. To be more precise, let $\{U_i\}_{i=1}^{\infty}$, $\{V_i\}_{i=1}^{\infty}$, be two renewal processes with common c.d.f. K and R respectively. By definition U_i and $U_j(i \neq j$ are independent and similarly for V_i and V_j. Concerning the relationship between the two processes assume only that $\{U_i + V_i\}_{i=1}^{\infty}$ forms a renewal process; that is, independence of U_i and V_i is not assumed. Let H denote the common c.d.f. of $U_i + V_i$ for all i. Define $T_0 = 0$ and, for $j \geqq 1$, set

$$T_{2j} = U_1 + V_1 + U_2 + V_2 + \cdots + U_j + V_j$$
$$T_{2j-1} = U_1 + V_1 + U_2 + V_2 + \cdots + U_j.$$

Define

$$A(t) = \begin{cases} 1 \text{ if } T_{2j-1} < t \leqq T_{2j} \text{ for some } j > 1 \\ 0 \quad \text{otherwise} \end{cases}$$

and

$$P_0(t) = 1 - P_1(t) = \Pr[A(t) = 0].$$

THEOREM 1. *For all* $t \geqq 0$

$$P_0(t) = \int_{0-}^{t} [1 - K(t - x)]\, dN(x)$$

where $N(x) = \sum_{j=0}^{\infty} H_j(x)$ *and* H_j *is the* c.d.f. *of* T_{2j} i.e., *the* jth *convolution of* H. *Moreover*,

$$P_0 = \lim_{t\to\infty} P_0(t) = \frac{E(U)}{E(U) + E(V)}$$

whenever at least one term of the denominator is finite. P_0 *is interpreted as being zero when* $E(V) = \infty$ *and one when* $E(U) = \infty$.

PROOF. By definition,

$$\begin{aligned} P_0(t) &= \sum_{j=0}^{\infty} \Pr[T_{2j} \leqq t < T_{2j+1}] \\ &= \sum_{j=0}^{\infty} \int_{0-}^{t} \Pr[T_{2j} \leqq t < T_{2j} + U_{j+1} \mid T_{2j} = x]\, dH_j(x) \\ &= \sum_{j=0}^{\infty} \int_{0-}^{t} [1 - K(t - x)]\, dH_j(x) \\ &= \int_{0-}^{t} [1 - K(t - x)]\, dN(x) \end{aligned}$$

as required. Since we are working with an at most countable family of r.v.'s, the conditional probability argument used above and in proofs which follow is valid. The second statement of the theorem is an immediate application of a theorem of Smith ([7], Theorem 1) which we quote in a particular form for further reference.

THEOREM S: *If $k(x)$ is any bounded function, zero for negative argument, integrable, non-increasing in $(0, \infty)$ for which $k(x) \to 0$ as $x \to \infty$; if H is a non-negative non-lattice distribution function and*

$$N(x) = \sum_{n=0}^{\infty} H_n(x)$$

then

$$\lim_{t\to\infty} \int_{0-}^{t} k(t - x)\, dN(x) = \int_0^{\infty} k(x)\, dx \left\{ \int_{0-}^{\infty} y\, dH(t) \right\}^{-1}.$$

The right hand side is to be taken as zero whenever the denominator is infinite.

In connection with the last statement of Theorem 1, observe that $P_1(t)$ converges to the stated limit since the function $k(x) = 1 - K(x)$ satisfies the conditions of Theorem S. We mention also that the last statement of Theorem 1 is actually a special case of a result concerning semi-Markov processes, given by Smith ([8], cf. Theorem 5).

4. The N-Process of the Type I Model. Set $p_0 = 0 = r_0$,

$$p_k = \Pr[n_1 = k] = \Pr[n_j - n_{j-1} = k] \qquad (j, k \varepsilon I^+)$$

and

$$r_k = \Pr[n_j = k \text{ for some } j] \qquad (k \varepsilon I^+).$$

Moreover, define the corresponding generating functions, for $|s| < 1$,

$$P(s) = \sum_{k=1}^{\infty} p_k s^k, \qquad R(s) = \sum_{k=1}^{\infty} r_k s^k.$$

The N-Process may be considered as a sampling of the positive integers I^+; that is, $n_1 < n_2 < n_3 < \cdots$ and $\{n_j, j \geqq 1\} \subset I^+$. In this context, one may speak of the event E, "an integer is sampled." One may show that, in the terminology of Feller [9], this event is recurrent. Since, for all $k \varepsilon I^+$

$$r_k = p_k + \sum_{j=1}^{k-1} p_j r_{k-j}$$

one obtains directly the known relationships

$$P(s) = \frac{R(s)}{1 + R(s)}, \qquad R(s) = \frac{P(s)}{1 - P(s)}$$

Moreover, it is known that (cf. [9])

$$\lim_{k\to\infty} r_{mk} = \lim_{s\to 1-} \frac{(1 - s^m)P(s)}{1 - P(s)} = m/E(n_1) \tag{2}$$

where m is the g.c.d. of those indices n for which $p_n > 0$. The right hand side of (2) is to be interpreted as zero whenever $E(n_1)$, the 'mean recurrence time,' is infinite.

The probabilities p_k are readily computed from the relation

$$p_k = \Pr[S_{k-1} \leqq Y_0 < S_k] \qquad (k \varepsilon I^+).$$

They are given in

LEMMA 1. *For all $k \varepsilon I^+$*

$$p_k = \int_{0-}^{\infty} [F_{k-1}(y) - F_k(y)]\, dG(y).$$

Observe that the event E is a certain event. That is

$$\sum_{k=1}^{\infty} p_k = 1 - \lim_{n\to\infty} \int_{0-}^{\infty} F_n(y)\, dG(y) = 1$$

since $\lim_{n\to\infty} F_n(y) = 0$ for all $y \geqq 0$ if and only if $F(0) < 1$, a condition which has been assumed.

Define the r.v. N_y for $y > 0$ as the smallest index k for which $S_k > y$. Set $Q_y(s)$ as the generating function of the probabilities associated with N_y. One may then easily show that for $|s| < 1$

$$P(s) = \int_{0-}^{\infty} Q_y(s)\, dG(y) = (1 - s) \sum_{k=0}^{\infty} s^k \int_{0-}^{\infty} G(y-)\, dF_k(y)$$

Consequently, setting $M_k(y) = E(N_y^k)$, one obtains

$$E(n_1^k) = \int_{0-}^{\infty} M_k(y)\, dG(y) \qquad (k \varepsilon I^+).$$

In particular

$$(3) \qquad E(n_1) = \int_{0-}^{\infty} M_1(y)\, dG(y) = \int_{0-}^{\infty} [1 - G(y-)]\, dM(y).$$

It is well known, and easily proven that

$$M_1(y) = \sum_{j=0}^{\infty} F_j(y)$$

$M_1(y)$ will be used very frequently throughout this paper. We shall therefore drop the subscript and write $M(y) = M_1(y)$.

Set $\mu = E(X)$ and $\nu = E(Y)$. It is well known, (cf. Smith [7]) that if $\mu < \infty$, $M(y) = y/\mu + o(y)$ as $y \to \infty$. Thus if $\mu < \infty$, by (3) $E(n_1) < \infty$ if and only if $\nu < \infty$. Similarly, if $\mu = \infty$, then $M(y) = o(y)$ and, hence $E(n_1) < \infty$ whenever $\nu < \infty$. The case of $\mu = \infty = \nu$ is special and will not be studied here.

5. The Z-renewal process. In this section the c.d.f. of Z as well as its Laplace-Stieltjes transform will be obtained. Consider the notation

$$\varphi(s) = \int_0^{\infty} e^{-sx}\, dF(x). \qquad \psi(s) = \int_{0-}^{\infty} e^{-sx}\, dG(x)$$

$$\Phi(s) = \int_0^{\infty} e^{-sx}\, dH(x), \qquad \psi^*(s) = \int_0^{\infty} e^{-sx} G(x-)\, dM(x)$$

for all $s \geqq 0$. One then obtains

THEOREM 2. *For all* $z \geqq 0$, $s \varepsilon R$

$$H(z) = \int_0^z G(u-)[1 - F(z-u)]\, dM(u) \tag{4}$$

$$\Phi(s) = [1 - \varphi(s)]\psi^*(s). \tag{5}$$

PROOF. Clearly

$$H(z) = \Pr[Z \leqq z] = \sum_{k=1}^{\infty} \Pr[S_{k-1} \leqq Y < S_k \leqq z].$$

For $k \geqq 2$

$$\begin{aligned}\Pr[S_{k-1} \leqq Y < S_k \leqq z] &= \int_{0-}^{z} \int_0^y [F(z-u) - F(y-u)]\, dF_{k-1}(u)\, dG(y)\\ &= \int_{0-}^{z} [G(z) - G(u-)]F(z-u)\, dF_{k-1}(u) - \int_{0-}^{z} F_k(y)\, dG(y)\\ &= G(z)F_k(z) - \int_0^z G(u-)F(z-u)\, dF_{k-1}(u) - \int_{0-}^{z} F_k(y)\, dG(y)\\ &= \int_0^z G(u-)\, dF_k(u) - \int_0^z G(u-)F(z-u)\, dF_{k-1}(u).\end{aligned}$$

For $k = 1$

$$\Pr[Y < S_1 \leqq Z] = \int_0^z G(u-)\, dF_1(u)$$

and (4) follows by summation over k. To obtain (5) for $s > 0$, write

$$\begin{aligned}\frac{1}{s}\Phi(s) &= \int_0^\infty e^{-sz} H(z)\, dz\\ &= \int_0^\infty \int_{u-}^\infty e^{-sz} G(u-)\, dz\, dM(u) - \int_0^\infty \int_{u-}^\infty e^{-sz} F(z-u)G(u-)\, dz\, dM(u)\\ &= \frac{1}{s}\psi^*(s) - \frac{1}{s}\varphi(s)\psi^*(s)\end{aligned}$$

as required. At $s = 0$, Φ may properly be defined by $\Phi(0) = 1$. This follows by an application of an Abelian theorem to (5). That is, consider

$$\begin{aligned}\lim_{s \to 0+} [1 - \varphi(s)]\psi^*(s) &= \lim_{s \to 0+} \int_{0-}^\infty e^{-sx} G(x-)\, dM(x) \left\{\int_{0-}^\infty e^{-sx}\, dM(x)\right\}^{-1}\\ &= \lim_{x \to \infty} G(x) = 1\end{aligned}$$

since $M(x) \to \infty$ as $x \to \infty$.

Of particular importance to the counter problem is the expectation of the secondary renewal process. One obtains

THEOREM 3. *$E(Z) < \infty$, if and only if $\nu < \infty$ and $\mu < \infty$. Moreover*

$$E(Z) = \mu E(n_1) = \mu \int_{0-}^{\infty} M(y)\, dG(y). \tag{6}$$

PROOF. The first statement follows from the relationship

$$\max(Y_0, X_1) \leqq Z_1 \leqq Y_0 + X_{n_1} \qquad \text{(a.s.)}.$$

The second statement is a consequence of a well known result in Sequential Analysis, for by it

$$E(Z \mid Y_0 = y) = \mu E(N_y)$$

and (6) follows by integration with respect to $dG(y)$. $E(Z)$ is to be interpreted as infinity whenever ν or $E(n_1)$ is infinite. Of course, (6) could also be proven directly from Theorem 2 using (5).

Let $N(x)$ denote the expected number of partial sums of the Z-process less than or equal to x; that is

$$N(x) = \sum_{j=0}^{\infty} H_j(x).$$

Define the bias of the counter at time x by $B(x) = M(x)/N(x)$. Then as a consequence of Theorem 3 and a known asymptotic renewal theorem, one obtains

LEMMA 2. *If $\mu < \infty$, then*

$$\lim_{x \to \infty} B(x) = 1/E(n_1)$$

where the right hand side is to be interpreted as zero when $\nu = +\infty$.

It may be easily shown that this result is also valid for the Type II *and Albert and Nelson models.*

6. The distribution of free-time. Let $W = Z_1 - Y_0$ represent the length of time the counter is free during successive registrations. Denote its c.d.f. and L-S transform by K and k respectively. Clearly $E(W) = \mu E(n_1) - \nu$. Moreover,

$$K(x) = \int_{0-}^{\infty} \Pr[Z_1 \leqq x + y \mid Y = y]\, dG(y).$$

Under the condition $[Y = y]$, Z_1 has a c.d.f. given by (4), but with G degenerate at y, i.e., $G(u) = 1$ if $u \geqq y$ and $G(u) = 0$ otherwise. Therefore,

$$K(x) = \int_{0-}^{\infty} \int_{y}^{x+y} [1 - F(x + y - u)]\, dM(u)\, dG(y)$$

$$= 1 - \int_{0-}^{\infty} \int_{0-}^{y} [1 - F(x + y - u)]\, dM(u)\, dG(y).$$

It follows similarly that $k(s)$ is the expectation w.r.t. $dG(y)$ of the L-S transform of $Z_1 - y$ obtained under the condition $[Y = y]$. By (5) this is seen to be

$$k(s) = [1 - \varphi(s)] \int_{0-}^{\infty} \int_{y}^{\infty} e^{-s(x-y)}\, dM(x)\, dG(y). \tag{7}$$

According to its definition in section 1, $P_1(t)$ is the probability that the counter is free at time t. Setting U and V of section 3 equal to Y and W, we have as a consequence of Theorem 1, the following result: for all $t \geqq 0$

$$P_0(t) = \int_{0-}^{t} [1 - G(t - x)]\, dN(x) \tag{8}$$

where

$$N(x) = \sum_{j=0}^{\infty} H_j(x).$$

This formula differs from equation (26) of Takacs [5]. Moreover, in the limit

$$P_0 = \lim_{t\to\infty} P_0(t) = \nu/\mu E(n_1).$$

Let the L-S transform of $P_0(t)$ be denoted by

$$\pi(s) = \int_{0-}^{\infty} e^{-st}\, dP_0(t).$$

Then, by direct computation one obtains from (8)

$$\pi(s) = \frac{1 - \psi(s)}{1 - \Phi(s)}.$$

7. Examples of the Type I counter problem. (a) $F(x) = 1 - e^{-\lambda x}$: This is the well known Poisson input counter problem which with various assumptions on G has been studied by several authors. For arbitrary G, the problem was treated by Takacs [3]. Because of special properties possessed by the exponential distribution, this particular example may be (and indeed has been) solved in several different ways. In [6], Smith has shown that much of the essential simplicity of this case carries over in asymptotic considerations to a wider class of F which generate so-called quasi-Poisson processes. For the present example, $\mu = 1/\lambda$ and $M(x) = \lambda x + 1$ for $x \geqq 0$ and $M(x) = 0$ for $x < 0$. The formulae of the previous sections become

$$Q_\nu(s) = se^{-\nu\lambda(1-s)}$$

$$P(s) = s\psi(\lambda - \lambda s)$$

$$E(n_1) = \lambda\nu + 1$$

$$H(z) = \int_{0-}^{z} [1 - e^{-\lambda(z-y)}]\, dG(y)$$

$$\Phi(s) = \varphi(s)\psi(s) = \frac{\lambda\psi(s)}{\lambda + s}.$$

These last two results may, of course, be obtained immediately from the known characterization of the exponential distribution that truncation on the left does not change the form of the distribution function. This implies that $Z_1 = Y_0 + X$

where X is exponentially distributed and is independent of Y_0. Finally, for this example, we have

$$\pi(s) = \frac{(\lambda + s)[1 - \psi(s)]}{\lambda + s - \lambda\psi(s)}$$

(b) $Y = d$ (a.s.): This important oft-studied case is applicable to counters for which the deadtime is independent of the intensity or amplitude of the incoming radioactive pulses. For this case $G(x) = 0$ or 1 according as $x <$ or $\geqq d$, and the formulae of the previous sections become

$$p_k = F_{k-1}(d) - F_k(d)$$

$$E(n_1) = M(d)$$

$$H(z) = \begin{cases} 0 & \text{if } z \leqq d \\ \int_d^z [1 - F(z - u)]\, dM(u) & \text{if } z > d \end{cases}$$

$$\Phi(s) = [1 - \varphi(s)] \int_d^\infty e^{-sx}\, dM(x) = 1 - [1 - \varphi(s)] \int_{0-}^d e^{-sx}\, dM(x)$$

and

$$\pi(s) = (1 - e^{-sd}) \left\{ [1 - \phi(s)] \int_{0-}^d e^{-sx}\, dM(x) \right\}^{-1}$$

(c) $G(y) = 1 - e^{-y\beta}$: The above two cases have been studied previously, whereas, the present case has not, to this author's knowledge, as yet been considered. In a different context, (7) has been employed by Scarf [4] for G exponential. For this example, we have $\nu = 1/\beta$.

$$p_k = [\varphi(\beta)]^{k-1}[1 - \varphi(\beta)]$$

$$P(s) = \frac{s[1 - \varphi(\beta)]}{1 - s\varphi(\beta)}$$

$$E(n_1) = [1 - \varphi(\beta)]^{-1}$$

$$H(z) = F(z) - \int_{0-}^z e^{-\beta x}[1 - F(z - x)]\, dM(x)$$

$$\Phi(s) = \frac{\varphi(s) - \varphi(\beta + s)}{1 - \varphi(\beta + s)}$$

$$\pi(s) = \frac{s[1 - \varphi(\beta + s)]}{(\beta + s)[1 - \varphi(s)]}$$

and

$$k(s) = \frac{\beta[\varphi(s) - \varphi(\beta - s)]}{(\beta - s)[1 - \varphi(s)][1 - \varphi(\beta - s)]}.$$

8. The general Type II counter. This problem is a very difficult one to solve in general. Discussions of the general problem have been given by Takacs [5] and Pollaczek [10]. Certain particular cases have been studied in the literature in greater detail. For example: Poisson input and constant deadtime by Feller [2], Poisson input and general deadtime by Takacs [3] and, with a different approach, by Chernoff and Daly [11], and exponentially distributed deadtime by Takacs [5]. The same notation as that used for the Type I problem will be employed in this section, but with the corresponding definition of n_i, namely (1).

In this model, it is simpler to evaluate r_k than p_k, contrary to what was observed in the Type I problem. For $k \geqq 1$

$$r_k = \Pr[S_r + Y_r < S_k\ r = 0, 1, \cdots, k-1] = \underbrace{\int_0^\infty \int_0^\infty \cdots \int_0^\infty}_{k} \tag{9}$$
$$\cdot G(x_1 + x_2 + \cdots + x_{k-1}-) \cdots G(x_1-)\, dF(x_1) \cdots dF(x_k).$$

Therefore, by the same argument leading to (2), one obtains the relationship

$$E(n_1) = m/\lim_{k\to\infty} r_{mk} \tag{10}$$

where m is the g.c.d. of those integers n for which $p_n > 0$. If $X \leqq Y$ (a.s.) set $m = 0$ and $n_1 = \infty$ (a.s.). In all other cases $\Pr[X > Y] > 0$. However, since $p_1 = \Pr[X > Y]$, one obtains $m = 1$. That is to say, whenever $\Pr[X > Y] > 0$

$$E(n_1) = 1/\lim_{k\to\infty} r_k$$

With a knowledge of r_k, one is able to compute $E(n_1)$ and hence the expectation of the secondary renewal process.

As before, set $Z_0 = 0$ (a.s.) and $Z_i = S_{n_i} - S_{n_{i-1}}$. Clearly the Z_i's form a renewal process. The problem of deriving an explicit expression for H, the common c.d.f. of the Z-renewal process, is extremely difficult. However, it is possible to display an integral equation which formally, but not always in practice, determines H. In section 10, an example will be given for which the solution is readily attained from this integral equation whereas it is not easily derived by other methods. Takacs [5] has, for the Type II problem, obtained an integral equation in $N(t)$, the expected number of counts (partial sums of the Z-process) in $[0, t]$ for all $t \geqq 0$. These two representations are equivalent in the sense that H and N are uniquely determined one by the other. More precisely, for $s \geqq 0$, the relationship between H and N is given by

$$\int_{0-}^\infty e^{-st}\, dN(t) = \sum_{j=0}^\infty \int_{0-}^\infty e^{-st}\, dH_j(t) = \frac{1}{1 - \Phi(s)}. \tag{11}$$

THEOREM 4. *For all* $z \geqq 0$

$$H(z) = \int_0^z \int_0^{z-x} [1 - H(z - x - t)] G(x + t-)\, dN(t)\, dF(x) \tag{12}$$

and for $s > 0$

(13) $$\Phi(s) = \lambda(s)[1 + \lambda(s)]^{-1}$$

where

(14) $$\lambda(s) = \int_{0-}^{\infty}\int_{0}^{\infty} e^{-s(x+t)}G(x+ t-)\, dF(x)\, dN(t).$$

(Notice that because of (11), $\lambda(s) + 1$ is the L-S transform of N.)

PROOF. (12) is obtained as follows. $[Z_1 \leqq z]$ is the union of two disjoint events, A and B say, where

$$A = [Y_0 < X_1 \leqq z]$$

and

$$B = [0 \leqq Y_0 - X_1 < Z_j \leqq z - X_1 \text{ for some } j \geqq 1].$$

Clearly

$$\Pr(A) = \int_0^z G(x-)\, dF(x).$$

Under the condition, $[z > Y_0 = y \geqq x = X_1] = C$ say,

$$\Pr(B \mid C) = 1 - \sum_{j=0}^{\infty} \Pr[Z_j \leqq y - x, Z_{j+1} > z - x]$$

$$= 1 - \int_{0-}^{y-x} [1 - H(z - x - t)]\, dN(t)$$

$$= \int_{y-x}^{z-x} [1 - H(z - x - t)]\, dN(t).$$

Therefore,

$$H(z) = \int_{0-}^{z} G(x-)\, dF(x) + \int_{0-}^{z-}\int_{x-}^{z-}\int_{y-x}^{z-x} [1 - H(z - x - t)]\, dN(t)\, dG(y)\, dF(x)$$

and an interchange of integration gives

$$H(z) = N(0)[F(z) - F(z-)]G(z-)$$

$$+ \int_{0-}^{z-}\int_{0-}^{z-x} [1 - H(z - x - t)]G(x + t-)\, dN(t)\, dF(x)$$

$$= \int_{0}^{z}\int_{0-}^{z-x} [1 - H(z - x - t)]G(x + t-)\, dN(t)\, dF(x)$$

as required. For the proof of (13), consider changes of integration according to

$$\int_0^{\infty} dz \int_0^z dx \int_{0-}^{z-x} dt = \int_0^{\infty} dx \int_{x-}^{\infty} dz \int_{0-}^{z-x} dt = \int_0^{\infty} dx \int_{0-}^{\infty} dt \int_{(x+t)-}^{\infty} dz.$$

It then follows that, for $x > 0$,

$$\frac{1}{s}\Phi(s) = \int_0^\infty e^{-sz} H(z)\, dz = \int_0^\infty \int_{0-}^\infty e^{-s(x+t)} \frac{1}{s}[1 - \Phi(s)]G(x + t-)\, dN(t)\, dF(x).$$

Solving for $\Phi(s)$ gives the desired result. As was stated earlier, Takacs [5] derived an integral equation in $N(t)$, which may be shortened to read

$$N(t) - 1 = \int_{0-}^{t} G(x-)\, dW(x)$$

where

$$W(x) = \int_{0-}^{x} F(x - y)\, dN(y).$$

Upon taking Laplace transforms of both sides one may check that it satisfies the relationship (11).

Of particular interest is the expectation of the secondary renewal process, namely $E(Z)$. As in the Type I problem, it follows from known results of Sequential Analysis that

$$E(Z) = E\left(\sum_{j=1}^{n_1} X_j\right) = \mu E(n_1). \tag{15}$$

From the above theorem, one obtains

$$E(Z) = \lim_{s\to 0} \frac{1 - \Phi(s)}{s} = 1/\lim_{s\to 0} s\lambda(s).$$

Thus, by (10), one obtains a double relationship

$$1/E(n_1) = \nu \lim_{s\to 0} s\lambda(s) = \lim_{k\to\infty} r_k\,.$$

Although it may well be that in a particular example one of the above limits will be computable, in most cases they will be unwieldy. For example, even in the case of Poisson input, the quantities r_k are complicated expressions, although $E(n_1)$ is a simple expression best obtained in an entirely different way using the particular properties of the exponential distribution. The p_k's may be expressed in terms of the r_j's as follows;

$$p_n = -\sum{}^{*} \prod_{j=1}^{n} \left((-r_j)^{k_j} \frac{1}{k_j\,!}\right) k.\,! \tag{16}$$

where $k. = \sum_{j=1}^{n} k_j$ and $\sum^*$ denotes summation over all vectors of integers $(k_1, k_2, \cdots, k_n)$ for which $\sum_{j=1}^{n} jk_j = n$. However, (16) will be, in most cases, very unwieldy, especially when one recalls the complicated structure of the r_j's.

9. The case of constant deadtime. Partial results for this example have been given by Takacs [5] for the Albert and Nelson model to be studied in the next

section. Also, this case has been studied from a different viewpoint by Smith [6]. We shall study this special case in full. Set $Y = d$ (a.s.). Then $G(x) = 0$ or 1 according as $x <$ or $\geqq d$. From (9) we obtain for $k \geqq 1$

$$r_k = 1 - F(d) \equiv q \text{ say.}$$

Consequently by (10)

$$E(n_1) = [1 - F(d)]^{-1} = q^{-1} \tag{17}$$

which is interpreted as being equal to ∞ if $1 = F(d)$. Using the notation introduced in section 4, one obtains

$$R(s) = \sum_{k=1}^{\infty} r_k s^k = qs(1 - s)^{-1}$$

and hence

$$P(s) = \frac{sq}{sq + 1 - s}.$$

From this relationship, or by direct computation, one obtains

$$p_k = q(1 - q)^{k-1}.$$

Therefore, n_1 has a Pascal (or geometric) distribution. The quickest way to obtain H and Φ for this example is as follows. Clearly $H(z) = 0$ for $z \leqq d$. For $z \geqq d$

$$H(z) = \sum_{n=1}^{\infty} \Pr [S_n \leqq z \mid n_1 = n] \, p_n$$

$$= q \sum_{n=1}^{\infty} \Pr [S_n \leqq z \mid n_1 = n](1 - q)^{n-1}.$$

Now

$$\Pr [S_n \leqq z \mid n_1 = n] = \Pr [S_n \leqq z \mid X_j < d, 1 \leqq j < n - 1, X_n > d]$$

$$= \Pr [U_1 + U_2 + \cdots + U_{n-1} + V \leqq z]$$

where the U_i's and V, are mutually independent with c.d.f.'s given by

$$\Pr [U_i \leqq u] \equiv K(u) = F(u)/F(d); \quad (u \leqq d, 1 \leqq i < n)$$

$$\Pr [V \leqq u] \equiv L(u) = \frac{F(u) - F(d)}{1 - F(d)} \qquad (u \geqq d).$$

Therefore, for $z \geqq$ d

$$H(z) = q \sum_{n=1}^{\infty} (1 - q)^{n-1} \int_{0-}^{z} K_n(z - u) \, dL(u) \tag{18}$$

where K_n denotes the nth convolution of K with itself. It is then immediate that

$$\Phi(s) = \frac{\int_d^\infty e^{-sx}\, dF(x)}{1 - \int_0^d e^{-sx}\, dF(x)}. \tag{19}$$

One may check that expressions (18) and (19) satisfy the equations of Theorem 4. In [6], Smith has obtained for this case $N(t) = 1$ for $0 \leqq t \leqq d$ and

$$N(t) - 1 = \int_{0-}^{t-d} [F(t - x) - F(d)]\, dM(x)$$

for $t > d$. By means of (11), one may show that this expression agrees with (19). (19) has also been obtained by Takacs [3]. From (15) and (17) it follows that $E(Z) = \mu q^{-1}$. One may also compute

$$\text{var}\,(Z) = q^{-1}\,\text{var}\,(X) + 2\mu q^{-2} \int_0^d x\, dF(x)$$

which disagrees with the expression given in Theorem 7 of [6].

For this example, not only is it possible to compute $P_0(t)$, the probability that the counter is free, but one may also derive the quantities $P_k(t)$ defined by

$$P_k(t) = \Pr\,[S_j + Y_j \geqq t \text{ for exactly } k \text{ values of } j]$$

for $k = 0, 1, 2, \cdots$. That is, $P_k(t)$ denotes the probability that k impulses are in process at time t. Now then

$$\begin{aligned} P_k(t) &= \sum_{j=0}^{\infty} \Pr\,[S_j \leqq t - d < S_{j+1} \leqq S_{j+k} < t \leqq S_{j+k+1}] \\ &= \int_{0-}^{t-d} \int_{t-d-x}^{t-x} [F_{k-1}(t - x - y) - F_k(t - x - y)]\, dF(y)\, dM(x). \end{aligned}$$

In particular

$$P_0(t) = \int_{0-}^{t-d} [1 - F(t - x)]\, dM(x). \tag{20}$$

Define the real functions h_m ($m \geqq 0$) as follows: for $v \leqq d$ set $h_m(v) = 1$ and for $v \geqq d$ set

$$h_m(v) = 1 - \int_0^{v-d} F_m(v - y)\, dF(y).$$

With these definitions we may write for $k \geqq 1$

$$P_k(t) = \int_{0-}^{t-d} [F_k(t - x) - F_{k+1}(t - x) - h_k(t - x) + h_{k-1}(t - x)]\, dM(x).$$

The functions h_m and $1 - F_m$ $(m \geqq 0)$ clearly satisfy the conditions of Theorem S, by which

$$(21) \qquad P_k = \lim_{t\to\infty} P_k(t) = \mu^{-1} \int_d^\infty [F_k(v) - F_{k+1}(v) - h_k(v) + h_{k-1}(v)]\, dv.$$

Moreover, by definition

$$\begin{aligned}\int_0^\infty [h_k(v) - h_{k-1}(v)]\, dv &= \int_d^\infty \int_0^{v-d} [F_{k-1}(v-y) - F_k(v-y)]\, dF(y)\, dv \\ &= \int_0^\infty \int_{y+d-}^\infty [F_{k-1}(v-y) - F_k(v-y)]\, dv\, dF(y) \\ &= \mu - \int_0^d [F_{k-1}(v) - F_k(v)]\, dv.\end{aligned}$$

Therefore, by (20) and (21)

$$P_k = \mu^{-1} \int_0^d [F_{k-1}(v) - 2F_k(v) + F_{k+1}(v)]\, dv \qquad (k \geqq 1)$$

$$P_0 = 1 - \mu^{-1} \int_0^d [1 - F(v)]\, dv$$

10. The Albert and Nelson generalization. Let $p \,\varepsilon\, [0, 1]$. Define

$$Y^{(p)} = \begin{cases} Y \text{ with probability } p \\ 0 \text{ with probability } 1 - p = q \end{cases}$$

which has c.d.f. G_p where $G_p(0) = p$, $G_p(x) = q + pG(x)$ for $x > 0$. Albert and Nelson [1] suggested as a generalization of the Types I and II counter models, the model in which the deadtime caused by an incoming pulse is Y or $Y^{(p)}$ according as the pulse is registered or not. Formally, define $n_0 = 0$ (a.s.) and $j \geqq 1$

$$(22) \qquad n_j = \min \{k \,\varepsilon\, I^+ : S_i + Y_i^{(p)} \leqq S_k\ (n_{j-1} < i < k),\ S_{n_{j-1}} + Y \leqq S_k\}$$

where as usual the subscript on $Y_i^{(p)}$ denotes identically and independent random variables with c.d.f. G_p. The purpose of this section is to show that the c.d.f. H of the secondary renewal process, $Z_j = S_{n_j} - S_{n_{j-1}}$ $(j \geqq 0)$ obtained for this generalization is in fact completely solved once the general Type II problem is solved, and in this sense this generalization is a very slight one.

Let $Z^{(p)}$ be the secondary renewal process of a Type II counter model in which the deadtime r.v. is $Y^{(p)}$. Let $H^{(p)}$ denote its c.d.f., $\Phi^{(p)}$ its characteristic function and $N^{(p)}(x) = \sum_{j=0}^\infty H_j^{(p)}(x)$. The distribution function of the Z-renewal process may then be given by

THEOREM 5. *For all* $z \geqq 0$

$$H(z) = \int_0^z \int_{0-}^{z-x} [1 - H^{(p)}(z - x - y)]\, G(x + y-)\, dN^{(p)}(y)\, dF(x)$$

and for $s > 0$

$$\Phi(s) = [1 - \Phi^{(p)}(s)] \int_{0-}^{\infty} \int_{0}^{\infty} e^{-s(x+y)} G(x + y-) \, dF(x) \, dN^{(p)}(y). \tag{23}$$

This theorem is proven in the same way as Theorem 4 upon noticing that in evaluating Pr $(B \mid C)$, one need only consider a Type II model with r.v.'s X and $Y^{(p)}$. Thus, although at first glance one might suspect that this more general model would offer difficulties peculiar to itself, it is seen that a solution of the corresponding Type II problem automatically provides a solution of the general problem. For $p = 1$, this model reduces to the Type II model and for $p = 0$, to the Type I model, as can be seen by comparing the definition of the N-process, (22), for these two values of p.

EXAMPLE. In [5], Takacs works out the special case of the Albert and Nelson model in which Y is constant a.e. As a further example, we evaluate here the case in which $G(x) = 1 - e^{-\lambda x}$ $(x \geqq 0)$. As was pointed out above, it will be sufficient to solve the Type II problem in which the deadtime, $Y^{(p)}$, has c.d.f. $G^{(p)}(x) = 1 - pe^{-\lambda x}$ $(x \geqq 0)$ and zero elsewhere. For this case we have by (14),

$$\begin{aligned}
\lambda^{(p)}(s) &= \int_{0-}^{\infty} \int_{0}^{\infty} e^{-s(x+y)} [1 - pe^{-\lambda(x+y)}] \, dF(x) \, dN^{(p)}(y) \\
&= \int_{0-}^{\infty} e^{-sy} [\varphi(s) - pe^{-(s+\lambda)y} \varphi(s + \lambda)] \, dN^{(p)}(y) \\
&= \frac{\varphi(s)}{1 - \Phi^{(p)}(s)} - p \frac{\varphi(s + \lambda)}{1 - \Phi^{(p)}(s + \lambda)}.
\end{aligned}$$

Therefore

$$\lambda^{(p)}(s) = \frac{\varphi(s) - q\varphi(s + \lambda)}{1 - \varphi(s)} - p \frac{\varphi(s + \lambda)}{1 - \varphi(s)} \lambda^{(p)}(s + \lambda)$$

since $1 - \Phi^{(p)}(s) = [1 + \lambda^{(p)}(s)]^{-1}$. Since this relation holds for all $s > 0$, we obtain by recursion that for all $n \geqq 1$

$$\begin{aligned}
\lambda^{(p)}(s) = \sum_{j=0}^{n} (-p)^j \frac{\varphi_j - p\varphi_{j+1}}{1 - \varphi_j} \prod_{k=0}^{j-1} \frac{\varphi_{k+1}}{1 - \varphi_k} \\
+ (-p)^{n+1} \prod_{k=0}^{n} \frac{\varphi_{k+1}}{1 - \varphi_k} \lambda^{(p)}(s + \lambda + n\lambda)
\end{aligned}$$

where for convenience we have set $\varphi_j = \varphi(s + j\lambda)$. Since $\varphi(s) \to 0$ and $\lambda(s) \to 0$ as $s \to \infty$ we finally obtain

$$\begin{aligned}
\lambda^{(p)}(s) &= \lim_{N \to \infty} \frac{1}{\varphi(s)} \sum_{j=0}^{N} (-p)^j \prod_{k=0}^{j} \frac{\varphi_k}{1 - \varphi_k} \\
&\quad + \sum_{j=1}^{N+1} (-p)^j \prod_{k=0}^{j} \frac{\varphi_{k+1}}{1 - \varphi_k} \\
&= \frac{1}{\varphi(s)} \sum_{j=0}^{\infty} (-p)^j \prod_{k=0}^{j} \frac{\varphi_k}{1 - \varphi_k}.
\end{aligned} \tag{24}$$

Thus $\Phi^{(p)}(s) = \lambda^{(p)}(s)[1 + \lambda^{(p)}(s)]^{-1}$, the solution to the Type II model in which the deadtime distribution is $G^{(p)}(x) = 1 - pe^{-\lambda x}$ $(x \geqq 0)$, is determined. From Theorem 5, in particular equation (23), one obtains

$$\Phi(s) = \varphi(s) - \varphi(s + \lambda) \frac{1 - \Phi^{(p)}(s)}{1 - \Phi^{(p)}(s + \lambda)}$$

$$= \varphi(s) - \varphi(s + \lambda) \frac{1 + \lambda^{(p)}(s + \lambda)}{1 + \lambda^{(p)}(s)}$$

Upon substitution of (24) into this expression, one obtains the solution to the Albert and Nelson model with exponential deadtime. When $p = 1$, (23) yields the solution to the corresponding Type II problem with exponential deadtime as given explicitly by Takacs [5] and implicitly by Pollaczek [10].

Acknowledgement. The author wishes to express his appreciation to Professor S. Karlin for many profitable discussions and suggestions, and to the referee for his helpful comments.

REFERENCES

[1] G. E. Albert and L. Nelson, "Contributions to the statistical theory of counter data," *Ann. Math. Stat.*, vol. 24 (1953), pp. 9–22.

[2] W. Feller, "On probability problems in the theory of counters," *Courant Anniversary Volume* (1948), pp. 105–15.

[3] L. Takacs, "On a probability problem arising in the theory of counters," *Proc. Cambridge Phil. Soc.*, vol. 52 (1956), pp. 488–498.

[4] K. J. Arrow, S. Karlin and H. Scarf, *Mathematical studies in inventory production processes*, to be published by Stanford University Press, (1958).

[5] L. Takacs, "On the sequence of events, selected by a counter from a recurrent process of events," *Akad Nauk. U.S.S.R., Probability Theory and its applications*, vol. 1 (1956), pp. 90–102.

[6] W. L. Smith, "On renewal theory, counter problems, and quasi-Poisson processes," *Proc. Cambridge Phil. Soc.*, vol. 53 (1957), pp. 175–193.

[7] W. L. Smith, "Asymptotic renewal theorems," *Proc. Roy. Soc. of Edinburgh*, Ser. A, vol. 64, pp. 9–48.

[8] W. L. Smith, "Regenerative stochastic processes," *Proc. Roy. Soc.*, Ser. A, vol. 232 (1955), pp. 6–31.

[9] W. Feller, *Introduction to probability theory and its applications*, vol. 1, J. Wiley and Sons, 1951.

[10] F. Pollaczek, "Sur la theorie stochastique des compteurs," *C. R. Acad. Sci. Paris*, vol. 238 (1954), pp. 766–7.

[11] H. Chernoff and J. F. Daly, "The distribution of shadows," *Journal of Math. and Mechanics*, vol. 6, No. 4 (1957), pp. 567–584.

MINIMUM VARIANCE UNBIASED ESTIMATION FOR THE TRUNCATED POISSON DISTRIBUTION

By R. F. Tate[1] and R. L. Goen[2]

University of Washington

1. Summary. A minimum variance unbiased estimator is provided for the parameter of a truncated Poisson distribution in the case of truncation on the left. In this connection the distribution is obtained for the sum of n independent identically distributed truncated Poisson random variables, and then well-known properties of sufficient statistics are employed to obtain the estimator. For the case of truncation away from the zero value results are expressed in terms of Stirling numbers of the second kind. The estimator has a particularly simple form and tables are available for its computation. For the general case results are expressed in terms of what we call generalized Stirling numbers. As a by-product of the statistical considerations there arises an identity between generalized Stirling numbers which may be useful in the study of Difference Equations.

2. Introduction. Numerous articles have been written on the subject of the estimation of the parameter of a truncated or censored Poisson distribution. Our work concerns the former distribution. The two types of distributions can be distinguished as follows: Consider an ordinary Poisson random variable with range $\{0, 1, 2, \cdots\}$, and let A be a subset of this range. If values in the set A cannot be members of a sample, then a random observation of the restricted variable is said to have a truncated Poisson distribution or to be truncated away from A. On the other hand there is the possibility for values in the set A to be members of a sample, but for some reason not distinguishable from one another. In this case a random observation of the restricted variable is said to have a censored Poisson distribution.

A situation calling for the truncated Poisson distribution would occur when one wishes to fit a distribution to Poisson-like data consisting of numbers of individuals in certain groups which possess a given attribute, but in which a group cannot be sampled unless at least a specified number of its members have the attribute. For example, the group may be a household of people, and the attribute measles; the specified number would then be one. A censored Poisson distribution is used most often in connection with pooled data.

The estimation problem for both the truncated and the censored cases has been discussed extensively from the point of view of maximum likelihood by Cohen [1]. Earlier results based on maximum likelihood were obtained by

---

Received July 3, 1956.

[1] Research Sponsored by ONR, Navy Theoretical Statistics Project.

[2] Submitted in partial satisfaction of the requirements for the degree of Master of Science in Mathematical Statistics.

Tippett [9], David and Johnson [2], and Rider [8]. Various other estimators were proposed by Moore [5], [6], Rider [8], and Plackett [7]. Plackett appears to be the only writer ever to propose an unbiased estimator for any of the cases of a truncated or censored Poisson distribution. His estimator for the parameter of a Poisson distribution truncated away from 0, which will arise several times during our discussion, is

$$\lambda^* = \frac{1}{n} \sum X_i ,$$

where the summation is taken over all $X_i \geq 2$.

The present paper is concerned with unbiased estimators for the case of tail truncation. It can readily be shown that truncation on the right, that is away from $A = \{c, c + 1, \cdots\}$, precludes the existence of an unbiased estimator. The argument is based on the identity of two power series; details will be omitted.

Assume that $A = \{0, 1, 2, \cdots, c\}$ for some $c \geqq 0$. Let the Poisson density be denoted by

$$f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \qquad x = 0, 1, 2, \cdots .$$

The density of the restricted random variable which is truncated away from A is then

$$g(x; \lambda, c) = \frac{e^{-\lambda} \lambda^x}{x! \sum_{c+1}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!}}, \qquad x = c + 1, c + 2, \cdots .$$

Consider a sample of n independent observations $X_1, X_2, \cdots, X_n$, each with density $g(x; \lambda, c)$, and let

$$T_c = \sum_{1}^{n} X_i .$$

It is well known that $\sum X_i$ is a sufficient statistic for the family $\{f(x; \lambda)\}$. A result of Tukey [10] states that sufficiency is preserved under truncation away from any Borel set in the range of X. Hence, in the case at hand T_c is sufficient for $\{g(x; \lambda, c)\}$. It can be verified that T_c is also complete.

For the case $c = 0$ the distribution of T_0 and the minimum variance unbiased estimator $\tilde{\lambda}_0$ are derived in Section 3. This is at the same time the most important case for applications and the easiest with which to deal. A recent extension of the table of Stirling numbers of the second kind makes $\tilde{\lambda}_0$ easy to compute for many values of n and T_0.

In order to express the results for the general case $c \geqq 1$ in a simple form it is necessary to introduce the notion of a generalized Stirling number. This will be done in Definition 3 below.

The following relations are quoted here for later reference.

Definition 1 (Jordan [3], p. 169): Stirling number of the second kind.

$$\mathfrak{S}_t^n = \frac{(-1)^n}{n!} \sum_0^\infty \binom{n}{k} (-1)^k (k)^t \qquad t = n, n+1, \cdots,$$

$\mathfrak{S}_t^n = 0$ for $t < n$.

Definition 2 (Jordan, p. 185):

$$\bar{C}_{p,i} = \sum_{j=p+1}^{2p-i} (-1)^{j+1} \binom{2p-1}{j} \mathfrak{S}_j^{j-p}.$$

Property 1 (Jordan, p. 169):

$$\mathfrak{S}_t^n = \mathfrak{S}_{t-1}^{n-1} + n\,\mathfrak{S}_{t-1}^n.$$

Property 2 (Jordan, p. 186):

$$\mathfrak{S}_{t+1}^{n+1} = \sum_{j=n}^{t} \binom{t}{j} \mathfrak{S}_j^n.$$

Property 3 (Jordan, p. 171):

$$\bar{C}_{t-n,t-2n} = n\,\bar{C}_{t-n-1,t-2n-1} + (t-1)\bar{C}_{t-n-1,t-2n}.$$

The generalized Stirling number will be introduced by

Definition 3:

$$\mathfrak{G}_{n,t}^c = \frac{(-1)^n t!}{n!} \sum \frac{n!}{k_1! \cdots k_{c+2}!} \frac{(-1)^{k_1} k_1^{(t - \sum_0^c j k_{j+2})}}{\left(t - \sum_0^c j k_{j+2}\right)! \prod_0^c (j!)^{k_{j+2}}},$$

where $k_i = 0, 1, \cdots, n$; $i = 1, 2, \cdots, c+2$; $t = n(c+1), n(c+1)+1, \cdots$; and the summation is taken over all $(k_1, \cdots, k_{c+2})$ such that $k_1 + \cdots + k_{c+2} = n$.

Property 4:

$$\mathfrak{G}_{n,t}^0 = \mathfrak{S}_t^n.$$

Property 5:

$$\mathfrak{G}_{n,t}^1 = \bar{C}_{t-n,t-2n}.$$

To verify Property 5 write $\mathfrak{G}_{n,t}^1$ as an iterated sum over k_1 and k_2, and use Definition 2.

In Section 4 the distribution of T_c and the minimum variance unbiased estimator $\bar{\lambda}_c$ are derived for the general case. There, also, a simple unbiased estimator based on one observation is given for λ, and is used, via the Lehmann-Scheffé-Blackwell method, to reproduce $\bar{\lambda}_c$. When equated, the two expressions for $\bar{\lambda}_c(t)$ provide an identity for the numbers $\mathfrak{G}_{n,t}^c$. The estimator used is related to Plackett's estimator λ^*.

3. The case $c = 0$. Let $X_1, X_2, \cdots, X_n$ be independent random variables,

each with density $g(x; \lambda, 0)$ and characteristic function $\phi_0(\alpha)$. Then T_0 has the characteristic function

$$\psi_0(\alpha) = [\phi_0(\alpha)]^n = \left(\sum_1^\infty \frac{\lambda^x e^{i\alpha x - \lambda}}{x!\,(1 - e^{-\lambda})}\right)^n.$$

Using the fact that $f(x; \lambda)$ has characteristic function $\exp\lambda(e^{i\alpha} - 1)$, and simplifying, we have

$$\psi_0(\alpha) = \left(\frac{e^{\lambda e^{i\alpha}} - 1}{e^\lambda - 1}\right)^n.$$

The inversion formula for characteristic functions shows that T_0 has the density

$$p_0(t) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \psi_0(\alpha)e^{-i\alpha t}\, d\alpha.$$

A binomial expansion for the numerator of $\psi_0(\alpha)$ shows that $p_0(t)$ is

$$\frac{(-1)^n}{(e^\lambda - 1)^n}\sum_{k=0}^n \binom{n}{k}(-1)^k \int_{-\pi}^{+\pi} \frac{e^{k\lambda e^{i\alpha} - i\alpha t}}{2\pi}\, d\alpha.$$

Since inversion of $\exp\lambda(e^{i\alpha} - 1)$ results in $e^{-\lambda}\lambda^t/t!$, the integral in $p_0(t)$ is $(k\lambda)^t/t!$, and from Definition 1 we finally arrive at

$$p_0(t) = \frac{\lambda^t n!}{(e^\lambda - 1)^n t!}\mathfrak{S}_t^n, \qquad t = n, n+1, \cdots.$$

It was noted in the introduction that T_0 is a complete sufficient statistic for the family $\{g(x; \lambda, 0)\}$. It then follows that if an unbiased estimator based on T_0 exists for λ, it will be unique and have the property of minimum variance (See Lehmann [4]). The condition for unbiasedness of $\tilde{\lambda}_0$ is

$$\sum_{t=n}^\infty \tilde{\lambda}_0(t)\frac{\lambda^t n!}{t!(e^\lambda - 1)^n}\mathfrak{S}_t^n \equiv \lambda.$$

In view of that fact that

$$(e^\lambda - 1)^n = \sum_{t=n}^\infty \frac{\lambda^t n!}{t!}\mathfrak{S}_t^n,$$

the condition becomes

$$\sum_{t=n}^\infty \tilde{\lambda}_0(t)\frac{\lambda^t}{t!}\mathfrak{S}_t^n \equiv \sum_{t=n}^\infty \frac{\lambda^{t+1}}{t!}\mathfrak{S}_t^n.$$

Comparing coefficients of powers of λ, we have the minimum variance unbiased estimator

$$\tilde{\lambda}_0(t) = t\,\frac{\mathfrak{S}_{t-1}^n}{\mathfrak{S}_t^n}.$$

Property 1 gives the alternative form[3]

$$\tilde{\lambda}_0(t) = \frac{t}{n}\left(1 - \frac{\mathfrak{S}_{t-1}^{n-1}}{\mathfrak{S}_t^n}\right).$$

Mr. Francis L. Miksa has computed the most complete table to date of Stirling numbers of the second kind.[4] Miksa's table gives $\mathfrak{S}_t^n$ for $n = 1(1)t$, $t = 1(1)50$. The quantity needed for the estimation of λ, the parameter of a Poisson distribution truncated away from zero, is

$$\tilde{\lambda}_0(t) = \frac{t}{n}\left(1 - \frac{\mathfrak{S}_{t-1}^{n-1}}{\mathfrak{S}_t^n}\right) = \frac{t}{n}\,C(n, t).$$

A table of $C(n, t)$ for $n = 2(1)t - 1$, $t = 3(1)50$ appears at the end of this paper. Note that for certain values of (n, t), $C(n, t)$ has not been tabulated, since

$$C(n, t) = 0 \quad \text{when} \quad n = t \geqq 1$$

$$C(1, t) = 1 \quad \text{when} \quad t \geqq 2.$$

All other missing entries are 1 (correct to 5 decimals); for example, $C(2, t) = 1$ for $t \geqq 19$.

For values of t which are large compared to n, the asymptotic expression $\mathfrak{S}_t^n \sim n^t/n!$ is available (Jordan [3], p. 173). Thus we have

$$\tilde{\lambda}_0(t) = \frac{t}{n}\left(1 - \left(\frac{n-1}{n}\right)^{t-1}\right).$$

The percentage error of approximation, $E(n, t)$, decreases with increasing t when n is fixed. For fixed t/n the percentage error increases with increasing n; however, the rate of increase falls off rapidly, as can be seen from the following short table, computed for $t/n = 4$.

(n, t)	(2, 8)	(4, 16)	(6, 24)	(8, 32)	(10, 40)	(12, 48)
$E(n, t)$	.006%	.046%	.073%	.090%	101%	.107%

Since $E(15, 50) = .4\%$, we may consider the approximation quite satisfactory for $2 \leqq n \leqq 15$, $t \geqq 51$. For larger values of n we must have t/n larger than 50/15, but not necessarily much larger, in view of the above table. For larger values of n one may also resort to the use of the unbiased estimator of Plackett, which was defined in the introduction and can also be written[5]

$$\lambda^* = \frac{t}{n}\left(1 - \frac{n_1}{t}\right),$$

[3] This form may be thought of as a slight change of the usual estimator t/n due to t missing zero class.

[4] This table is as yet unpublished

[5] In this connection see also the definition of V_c in Section 4

here n_1 is the number of observations in the sample which have the value 1; r one can use the maximum likelihood estimator $\hat{\lambda}$, which is the solution of the quation

$$\frac{t}{n} = \frac{\hat{\lambda}}{1 - e^{-\hat{\lambda}}}.$$

In summary, the following is an improved procedure for estimating λ, which ı many cases yields a minimum variance unbiased estimator.

Estimate λ by

$$\frac{t}{n} C(n, t), \qquad 1 \leqq n \leqq t, 1 \leqq t \leqq 50; n = t \geqq 51; n = 1, t \geqq 51,$$

$$\frac{t}{n}\left(1 - \left(\frac{n-1}{n}\right)^{t-1}\right), \qquad 2 \leqq n \leqq 15, t \geqq 51; n \geqq 16, t \gg n,$$

$$\lambda^* \quad \text{or} \quad \hat{\lambda}, \qquad \text{otherwise.}$$

An extended table of $C(n, t)$ would be quite useful. However, in order to obtain ch a table it is necessary to devise a method for computing $C(n, t)$ which does ɔt depend on entries in the table of $\mathfrak{S}_t^n$, since, for example, $\mathfrak{S}_{50}^{16}$ in Miksa's ble is an integer of forty-seven digits. The authors have been unable to do this.

The following facts should be observed in comparing our estimator with the timator λ^* of Plackett [7] and the maximum likelihood estimator $\hat{\lambda}$.

1. Plackett's estimator λ^* and $\tilde{\lambda}_0$ are different, and λ^* has exact variance

$$\frac{1}{n}\left(\lambda + \frac{\lambda^2}{e^{\lambda} - 1}\right).$$

2. $\hat{\lambda}$ was shown by David and Johnson [2] to be the solution of the equation

$$\bar{x} = \frac{\lambda}{1 - e^{-\lambda}},$$

it is obviously a function of T_0. A simple numerical calculation shows that and $\tilde{\lambda}_0$ are different. Therefore, by uniqueness of unbiased estimators based . T_0, we see that $\hat{\lambda}$ is a biased estimator of λ.

3. David and Johnson also showed that the asymptotic variance of $\hat{\lambda}$ is

$$\frac{\lambda(1 - e^{-\lambda})^2}{n(1 - e^{-\lambda} - \lambda e^{-\lambda})}.$$

ıis is then also, for each fixed n, the Cramér-Rao lower bound for exact variices of unbiased estimators. The following calculations show that there is no biased estimator whose variance attains this lower bound: Let $J_\lambda(\mathbf{x})$ denote ɜ joint density of n independent truncated Poisson random variables, each th density $g(x; \lambda, 0)$. Then, a necessary and sufficient condition for a Cramér-,o estimator to exist is that there exist a function $h(\lambda)$ such that the expression

$$\lambda + h(\lambda) \frac{\partial \log J_\lambda(\mathbf{x})}{\partial \lambda}$$

is independent of λ for all values of $\mathbf{x}$. It can be verified that

$$\frac{\partial \log J_\lambda(\mathbf{x})}{\partial \lambda} = \left(\frac{\sum x_k}{n} - \frac{n}{1 - e^{-\lambda}} \right),$$

and that no such function $h(\lambda)$ exists. Moreover, since $\bar{\lambda}_0$ and λ^* are differe functions of t, the variance of λ^* must exceed that of $\bar{\lambda}_0$. Consequently, we ma write (for each fixed n)

$$\frac{\lambda(1 - e^{-\lambda})^2}{n(1 - e^{-\lambda} - \lambda e^{-\lambda})} < \sigma^2_{\bar{\lambda}_0} < \frac{1}{n}\left(\lambda + \frac{\lambda^2}{e^\lambda - 1}\right).$$

4. The general case. The derivation of the distribution of T_c and the minimu variance unbiased estimator $\bar{\lambda}_c$ proceeds in a manner analogous to the ca $c = 0$, except that here we use a multinomial expansion and generalized Stirli numbers. More precisely, let

$$F(c) = \sum_{j=0}^{c} \frac{e^{-\lambda}\lambda^j}{j!}.$$

Then, T_c will have characteristic function

$$\psi_c(\alpha) = \left(\sum_{c+1}^{\infty} \frac{e^{i\alpha x - \lambda}\lambda^x}{x!\,[1 - F(c)]} \right)^n = \frac{e^{-n\lambda}}{[1 - F(c)]^n} \left(e^{\lambda e^{i\alpha}} - \sum_{0}^{c} \frac{e^{i\alpha x}\lambda^x}{x!} \right)^n.$$

After performing the multinomial expansion, employing the inversion formul and evaluating the same types of integrals as before, we use Definition 3 a arrive at the following expression for the density of T_c:

$$p_c(t) = \frac{n!\,\lambda^t \mathfrak{S}^c_{n,t}}{t!\left(e^\lambda - \sum_0^c \frac{\lambda^j}{j!}\right)^n}, \qquad t = n(c+1), n(c+1)+1, \cdots$$

In the same way as before the condition of unbiasedness now yields

$$\bar{\lambda}_c(t) = t\,\frac{\mathfrak{S}^c_{n,t-1}}{\mathfrak{S}^c_{n,t}}.$$

It is clear from Property 4 that for $c = 0$, $\bar{\lambda}_c(t)$ reduces to the expression of Se tion 3 Also, from Property 5 we see that

$$\bar{\lambda}_1(t) = t\,\frac{\bar{C}_{t-n-1,t-2n-1}}{\bar{C}_{t-n,t-2n}}.$$

At the present time there appears to be only one available table (Jordan, p. 17 for evaluating $\bar{C}_{p,s}$. This table handles the estimation problem for $n = 1, \cdots,$ $2n + 1 \leqq t \leqq n + 6$.

One simple unbiased estimator for λ is

$$U_c(x_1) = \begin{cases} 0 & x_1 = c + 1 \\ x_1 & x_1 \geqq c + 2. \end{cases}$$

Now we use the Lehmann-Scheffé-Blackwell Method (see Lehmann [4]):

$$\tilde{\lambda}_c(t) = E(U_c \mid T_c = t) = \sum_{c+2}^{t-(n-1)(c+1)} xP(X_1 = x \mid T_c = t).$$

$P(X_1 = x \mid T_c = t)$ can be written as $P(X_1 = x)\, P(\sum_2^n X_j = t - x)/P(T_c = t)$, and then simplified by the use of $p_c(t)$ to the form

$$P(X_1 = x \mid T_c = t) = \frac{\binom{t}{x} \mathfrak{S}_{n-1,t-x}^c}{n\mathfrak{S}_{n,t}^c}.$$

Substituting this expression in the above, and equating the result with the earlier form of $\tilde{\lambda}_c(t)$, we obtain the identity

$$\sum_{(c+1)(n-1)}^{t-c-2} \binom{t-1}{j} \mathfrak{S}_{n-1,j}^c = n\mathfrak{S}_{n,t-1}^c .$$

For $c = 0$ this reduces to a combination of Property 1 and Property 2.

The natural unbiased estimator based on the whole sample, which may be generated from U_c, is

$$V_c(x_1, x_2, \cdots, x_n) = \frac{1}{n} \sum_1^n U_c(x_j).$$

V_0 is precisely Plackett's λ^*.

TABLE OF $10^5\, C(n, t) = 10^5 \left(1 - \frac{\mathfrak{S}_{t-1}^{n-1}}{\mathfrak{S}_t^n}\right)$

(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$
(2, 3)	66667	(3, 7)	89701	(3, 27)	99997	(4, 19)	99428	(4, 39)	99998	(5, 20)	98497	(5, 40)	99983
(2, 4)	85714	(3, 8)	93478	(3, 28)	99998	(4, 20)	99572	(4, 40)	99998	(5, 21)	98807	(5, 41)	99987
(2, 5)	93333	(3, 9)	95802	(3, 29)	99999	(4, 21)	99680	(4, 41)	99999	(5, 22)	99052	(5, 42)	99989
(2, 6)	96774	(3, 10)	97267	(3, 30)	99999	(4, 22)	99761	(4, 42)	99999	(5, 23)	99245	(5, 43)	99991
(2, 7)	98413	(3, 11)	98207	(3, 31)	99999	(4, 23)	99821	(4, 43)	99999	(5, 24)	99399	(5, 44)	99993
(2, 8)	99213	(3, 12)	98818	$n = 4$		(4, 24)	99866	$n = 5$		(5, 25)	99521	(5, 45)	99995
(2, 9)	99608	(3, 13)	99218	(4, 5)	40000	(41 25)	99898	(5, 6)	33333	(5, 26)	99618	(5, 46)	99996
(2, 10)	99804	(3, 14)	99481	(4, 6)	61538	(4, 26)	99925	(5, 7)	53571	(5, 27)	99695	(5, 47)	99997
(2, 11)	99902	(3, 15)	99655	(4, 7)	74286	(4, 27)	99943	(5, 8)	66667	(5, 28)	99756	(5, 48)	99997
(2, 12)	99951	(3, 16)	99771	(4, 8)	82305	(4, 28)	99958	(5, 9)	75529	(5, 29)	99805	(5, 49)	99998
(2, 13)	99976	(3, 17)	99847	(4, 9)	87568	(4, 29)	99968	(5, 10)	81728	(5, 30)	99844	(5, 50)	99998
(2, 14)	99988	(3, 18)	99898	(4, 10)	91130	(4, 30)	99976	(5, 11)	86177	(5, 31)	99876	$n = 6$	
(2, 15)	99994	(3, 19)	99932	(4, 11)	93586	(4, 31)	99982	(5, 12)	89434	(5, 32)	99901	(6, 7)	28571
(2, 16)	99997	(3, 20)	99955	(4, 12)	95339	(4, 32)	99987	(5, 13)	91851	(5, 33)	99921	(6, 8)	47368
(2, 17)	99998	(3, 21)	99970	(4, 13)	96583	(4, 33)	99990	(5, 14)	93686	(5, 34)	99936	(6, 9)	60317
(2, 18)	99999	(3, 22)	99980	(4, 14)	97482	(4, 34)	99992	(5, 15)	95070	(5, 35)	99949	(6, 10)	69549
$n = 3$		(3, 23)	99987	(4, 15)	98137	(4, 35)	99994	(5, 16)	96136	(5, 36)	99959	(6, 11)	76307
(3, 4)	50000	(3, 24)	99991	(4, 16)	98618	(4, 36)	99996	(5, 17)	96961	(5, 37)	99968	(6, 12)	81360
(3, 5)	72000	(3, 25)	99994	(4, 17)	98971	(4, 37)	99997	(5, 18)	97602	(5, 38)	99974	(6, 13)	85202
(3, 6)	83333	(3, 26)	99996	(4, 18)	99233	(4, 38)	99998	(5, 19)	98104	(5, 39)	99979	(6, 14)	88164

TABLE—*Continued*

(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$
(6, 15)	90474	(7, 34)	99368	(9, 12)	46666	(10, 34)	96576	(12, 18)	59241	(13, 43)	96100	(15, 33)	84943
(6, 16)	92294	(7, 35)	99460	(9, 13)	55765	(10, 35)	96946	(12, 19)	64243	(13, 44)	96427	(15, 34)	86196
(6, 17)	93738	(7, 36)	99538	(9, 14)	63008	(10, 36)	97273	(12, 20)	68514	(13, 45)	96726	(15, 35)	87330
(6, 18)	94893	(7, 37)	99605	(9, 15)	68847	(10, 37)	97564	(12, 21)	72182	(13, 46)	96998	(15, 36)	88360
(6, 19)	95822	(7, 38)	99662	(9, 16)	73607	(10, 38)	97822	(12, 22)	75348	(13, 47)	97246	(15, 37)	89297
(6, 20)	96573	(7, 39)	99711	(9, 17)	77523	(10, 39)	98052	(12, 23)	78095	(13, 48)	97473	(15, 38)	90149
(6, 21)	97182	(7, 40)	99753	(9, 18)	80771	(10, 40)	98257	(12, 24)	80489	(13, 49)	97680	(15, 39)	90926
(6, 22)	97678	(7, 41)	99768	(9, 19)	83485	(10, 41)	98439	(12, 25)	82582	(13, 50)	97869	(15, 40)	91636
(6, 23)	98084	(7, 42)	99820	(9, 20)	85765	(10, 42)	98602	(12, 26)	84419	$n = 14$		(15, 41)	92285
(6, 24)	98417	(7, 43)	99845	(9, 21)	87693	(10, 43)	98747	(12, 27)	86036	(14, 15)	13333	(15, 42)	92878
(6, 25)	98690	(7, 44)	99867	(9, 22)	89331	(10, 44)	98877	(12, 28)	87465	(14, 16)	24419	(15, 43)	93422
(6, 26)	98915	(7, 45)	99886	(9, 23)	90727	(10, 45)	98993	(12, 29)	88730	(14, 17)	33725	(15, 44)	93921
(6, 27)	99101	(7, 46)	99902	(9, 24)	91924	(10, 46)	99096	(12, 30)	89853	(14, 18)	41607	(15, 45)	94379
(6, 28)	99254	(7, 47)	99916	(9, 25)	92952	(10, 47)	99189	(12, 31)	90852	(14, 19)	48331	(15, 46)	94799
(6, 29)	99381	(7, 48)	99928	(9, 26)	93838	(10, 48)	99272	(12, 32)	91743	(14, 20)	54107	(15, 47)	95186
(6, 30)	99485	(7, 49)	99939	(9, 27)	94605	(10, 49)	99347	(12, 33)	92539	(14, 21)	59098	(15, 48)	95542
(6, 31)	99572	(7, 50)	99947	(9, 28)	95269	(10, 50)	99413	(12, 34)	93251	(14, 22)	63434	(15, 49)	95870
(6, 32)	99644	$n = 8$		(9, 29)	95846	$n = 11$		(12, 35)	93890	(14, 23)	67220	(15, 50)	96172
(6, 33)	99704	(8, 9)	22222	(9, 30)	96348	(11, 12)	16667	(12, 36)	94463	(14, 24)	70539	$n = 16$	
(6, 34)	99754	(8, 10)	38400	(9, 31)	96787	(11, 13)	29864	(12, 37)	94978	(14, 25)	73462	(16, 18)	21769
(6, 35)	99795	(8, 11)	50505	(9, 32)	97170	(11, 14)	40476	(12, 38)	95442	(14, 26)	76044	(16, 19)	30341
(6, 36)	99830	(8, 12)	59763	(9, 33)	97505	(11, 15)	49120	(12, 39)	95860	(14, 27)	78333	(16, 20)	37735
(6, 37)	99858	(8, 13)	66972	(9, 34)	97799	(11, 16)	56240	(12, 40)	96238	(14, 28)	80369	(16, 21)	44153
(6, 38)	99882	(8, 14)	72670	(9, 35)	98057	(11, 17)	62162	(12, 41)	96579	(14, 29)	82185	(16, 22)	49754
(6, 39)	99902	(8, 15)	77229	(9, 36)	98283	(11, 18)	67127	(12, 42)	96887	(14, 30)	83809	(16, 23)	54665
(6, 40)	99918	(8, 16)	80916	(9, 37)	98482	(11, 19)	71323	(12, 43)	97166	(14, 31)	85264	(16, 24)	58991
(6, 41)	99932	(8, 17)	83925	(9, 38)	98658	(11, 20)	74890	(12, 44)	97419	(14, 32)	86571	(16, 25)	62818
(6, 42)	99943	(8, 18)	86400	(9, 39)	98812	(11, 21)	77942	(12, 45)	97648	(14, 33)	87747	(16, 26)	66215
(6, 43)	99953	(8, 19)	88451	(9, 40)	98949	(11, 22)	80565	(12, 46)	97856	(14, 34)	88808	(16, 27)	69242
(6, 44)	99961	(8, 20)	90159	(9, 41)	99069	(11, 23)	82831	(12, 47)	98045	(14, 35)	89767	(16, 28)	71946
(6, 45)	99967	(8, 21)	91590	(9, 42)	99175	(11, 24)	84797	(12, 48)	98217	(14, 36)	90635	(16, 29)	74370
(6, 46)	99973	(8, 22)	92795	(9, 43)	99269	(11, 25)	86508	(12, 49)	98373	(14, 37)	91421	(16, 30)	76508
(6, 47)	99977	(8, 23)	93813	(9, 44)	99352	(11, 26)	88003	(12, 50)	98514	(14, 38)	92135	(16, 31)	78510
(6, 48)	99981	(8, 24)	94676	(9, 45)	99426	(11, 27)	89314	$n = 13$		(14, 39)	92783	(16, 32)	80281
(6, 49)	99984	(8, 25)	95411	(9, 46)	99491	(11, 28)	90466	(13, 14)	14286	(14, 40)	93374	(16, 33)	81884
(6, 50)	99987	(8, 26)	96037	(9, 47)	99548	(11, 29)	91481	(13, 15)	26000	(14, 41)	93912	(16, 34)	83337
$n = 7$		(8, 27)	96574	(9, 48)	99599	(11, 30)	92377	(13, 16)	35714	(14, 42)	94403	(16, 35)	84657
(7, 8)	25000	(8, 28)	97034	(9, 49)	99644	(11, 31)	93171	(13, 17)	43849	(14, 43)	95851	(16, 36)	85859
(7, 9)	42424	(8, 29)	97429	(9, 50)	99684	(11, 32)	93875	(13, 18)	50719	(14, 44)	95260	(16, 37)	86954
(7, 10)	55000	(8, 30)	97770	$n = 10$		(11, 33)	94501	(13, 19)	56565	(14, 45)	95635	(16, 38)	87953
(7, 11)	64326	(8, 31)	98063	(10, 11)	18182	(11, 34)	95058	(13, 20)	61573	(14, 46)	95978	(16, 39)	88867
(7, 12)	71392	(8, 32)	98317	(10, 12)	32258	(11, 35)	95555	(13, 21)	65888	(14, 47)	96293	(16, 40)	89703
(7, 13)	76841	(8, 33)	98536	(10, 13)	43357	(11, 36)	95999	(13, 22)	69627	(14, 48)	96581	(16, 41)	90469
(7, 14)	81104	(8, 34)	98726	(10, 14)	52242	(11, 37)	96395	(13, 23)	72881	(14, 49)	96846	(16, 42)	91173
(7, 15)	84480	(8, 35)	98890	(10, 15)	59447	(11, 38)	96750	(13, 24)	75726	(14, 50)	97059	(16, 43)	91819
(7, 16)	87181	(8, 36)	99033	(10, 16)	65354	(11, 39)	97068	(13, 25)	78224	$n = 15$		(16, 44)	92413
(7, 17)	89362	(8, 37)	99157	(10, 17)	70243	(11, 40)	97354	(13, 26)	80424	(15, 16)	12500	(16, 45)	92960
(7, 18)	91135	(8, 38)	99265	(10, 18)	74324	(11, 41)	97610	(13, 27)	82369	(15, 17)	23018	(16, 46)	93464
(7, 19)	92556	(8, 39)	99359	(10, 19)	77755	(11, 42)	97841	(13, 28)	84093	(15, 18)	31944	(16, 47)	93929
(7, 20)	93780	(8, 40)	99441	(10, 20)	80657	(11, 43)	98048	(13, 29)	85625	(15, 19)	39578	(16, 48)	94358
(7, 21)	94769	(8, 41)	99512	(10, 21)	83127	(11, 44)	98235	(13, 30)	86991	(15, 20)	46150	(16, 49)	94754
(7, 22)	95589	(8, 42)	99574	(10, 22)	85239	(11, 45)	98403	(13, 31)	88211	(15, 21)	51843	(16, 50)	95130
(7, 23)	96274	(8, 43)	99628	(10, 23)	87054	(11, 46)	98555	(13, 32)	89304	(15, 22)	56801	$n = 17$	
(7, 24)	96846	(8, 44)	99675	(10, 24)	88619	(11, 47)	98691	(13, 33)	90283	(15, 23)	61139	(17, 18)	11111
(7, 25)	97327	(8, 45)	99716	(10, 25)	89975	(11, 48)	98815	(13, 34)	91164	(15, 24)	64953	(17, 19)	20648
(7, 26)	97731	(8, 46)	99752	(10, 26)	91152	(11, 49)	98926	(13, 35)	91957	(15, 25)	68319	(17, 20)	28888
(7, 27)	98072	(8, 47)	99783	(10, 27)	92178	(11, 50)	99027	(13, 36)	92672	(15, 26)	71301	(17, 21)	36054
(7, 28)	98360	(8, 48)	99810	(10, 28)	93074	$n = 12$		(13, 37)	93318	(15, 27)	73951	(17, 22)	42317
(7, 29)	98603	(8, 49)	99834	(10, 29)	93859	(12, 13)	15385	(13, 38)	93902	(15, 28)	76314	(17, 23)	47620
(7, 30)	98810	(8, 50)	99855	(10, 30)	94548	(12, 14)	27799	(13, 39)	94430	(15, 29)	78428	(17, 24)	52678
(7, 31)	98985	$n = 9$		(10, 31)	95154	(12, 15)	37949	(13, 40)	94910	(15, 30)	80322	(17, 25)	56979
(7, 32)	99134	(9, 10)	20000	(10, 32)	95688	(12, 16)	46340	(13, 41)	95345	(15, 31)	82025	(17, 26)	60807
(7, 33)	99260	(9, 11)	35065	(10, 33)	96159	(12, 17)	53344	(13, 42)	95740	(15, 32)	83559	(17, 27)	64222

One simple unbiased estimator for λ is

$$U_c(x_1) = \begin{cases} 0 & x_1 = c + 1 \\ x_1 & x_1 \geqq c + 2. \end{cases}$$

Now we use the Lehmann-Scheffé-Blackwell Method (see Lehmann [4]):

$$\tilde{\lambda}_c(t) = E(U_c \mid T_c = t) = \sum_{c+2}^{t-(n-1)(c+1)} xP(X_1 = x \mid T_c = t).$$

$P(X_1 = x \mid T_c = t)$ can be written as $P(X_1 = x)\, P(\sum_2^n X_j = t - x)/P(T_c = t)$, and then simplified by the use of $p_c(t)$ to the form

$$P(X_1 = x \mid T_c = t) = \frac{\binom{t}{x} \mathfrak{S}^c_{n-1,t-x}}{n\mathfrak{S}^c_{n,t}}.$$

Substituting this expression in the above, and equating the result with the earlier form of $\tilde{\lambda}_c(t)$, we obtain the identity

$$\sum_{(c+1)(n-1)}^{t-c-2} \binom{t-1}{j} \mathfrak{S}^c_{n-1,j} = n\mathfrak{S}^c_{n,t-1}.$$

For $c = 0$ this reduces to a combination of Property 1 and Property 2.

The natural unbiased estimator based on the whole sample, which may be generated from U_c, is

$$V_c(x_1, x_2, \cdots, x_n) = \frac{1}{n} \sum_1^n U_c(x_j).$$

V_0 is precisely Plackett's λ^*.

TABLE OF $10^5\, C(n, t) = 10^5 \left(1 - \dfrac{\mathfrak{S}^{n-1}_{t-1}}{\mathfrak{S}^n_t}\right)$

(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$
(2, 3)	66667	(3, 7)	89701	(3, 27)	99997	(4, 19)	99428	(4, 39)	99998	(5, 20)	98497	(5, 40)	99983
(2, 4)	85714	(3, 8)	93478	(3, 28)	99998	(4, 20)	99572	(4, 40)	99998	(5, 21)	98807	(5, 41)	99987
(2, 5)	93333	(3, 9)	95802	(3, 29)	99999	(4, 21)	99680	(4, 41)	99999	(5, 22)	99052	(5, 42)	99989
(2, 6)	96774	(3, 10)	97267	(3, 30)	99999	(4, 22)	99761	(4, 42)	99999	(5, 23)	99245	(5, 43)	99991
(2, 7)	98413	(3, 11)	98207	(3, 31)	99999	(4, 23)	99821	(4, 43)	99999	(5, 24)	99399	(5, 44)	99993
(2, 8)	99213	(3, 12)	98818	$n = 4$		(4, 24)	99866	$n = 5$		(5, 25)	99521	(5, 45)	99995
(2, 9)	99608	(3, 13)	99218	(4, 5)	40000	(4, 25)	99898	(5, 6)	33333	(5, 26)	99618	(5, 46)	99996
(2, 10)	99804	(3, 14)	99481	(4, 6)	61538	(4, 26)	99925	(5, 7)	53571	(5, 27)	99695	(5, 47)	99997
(2, 11)	99902	(3, 15)	99655	(4, 7)	74286	(4, 27)	99943	(5, 8)	66667	(5, 28)	99756	(5, 48)	99997
(2, 12)	99951	(3, 16)	99771	(4, 8)	82305	(4, 28)	99958	(5, 9)	75529	(5, 29)	99805	(5, 49)	99998
(2, 13)	99976	(3, 17)	99847	(4, 9)	87508	(4, 29)	99968	(5, 10)	81728	(5, 30)	99844	(5, 50)	99998
(2, 14)	99988	(3, 18)	99898	(4, 10)	91130	(4, 30)	99976	(5, 11)	86177	(5, 31)	99876	$n = 6$	
(2, 15)	99994	(3, 19)	99932	(4, 11)	93586	(4, 31)	99982	(5, 12)	89434	(5, 32)	99901	(6, 7)	28571
(2, 16)	99997	(3, 20)	99955	(4, 12)	95339	(4, 32)	99987	(5, 13)	91851	(5, 33)	99921	(6, 8)	47368
(2, 17)	99998	(3, 21)	99970	(4, 13)	96583	(4, 33)	99990	(5, 14)	93686	(5, 34)	99936	(6, 9)	60317
(2, 18)	99999	(3, 22)	99980	(4, 14)	97482	(4, 34)	99992	(5, 15)	95070	(5, 35)	99949	(6, 10)	69549
$n = 3$		(3, 23)	99987	(4, 15)	98137	(4, 35)	99994	(5, 16)	96136	(5, 36)	99959	(6, 11)	76307
(3, 4)	50000	(3, 24)	99991	(4, 16)	98618	(4, 36)	99996	(5, 17)	96961	(5, 37)	99968	(6, 12)	81360
(3, 5)	72000	(3, 25)	99994	(4, 17)	98971	(4, 37)	99997	(5, 18)	97602	(5, 38)	99974	(6, 13)	85202
(3, 6)	83333	(3, 26)	99996	(4, 18)	99233	(4, 38)	99998	(5, 19)	98101	(5, 39)	99979	(6, 14)	88164

TABLE—*Continued*

(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$
(6, 15)	90474	(7, 34)	99368	(9, 12)	46666	(10, 34)	96576	(12, 18)	59241	(13, 43)	96100	(15, 33)	84943
(6, 16)	92294	(7, 35)	99460	(9, 13)	55765	(10, 35)	96946	(12, 19)	64246	(13, 44)	96427	(15, 34)	86196
(6, 17)	93738	(7, 36)	99538	(9, 14)	63008	(10, 36)	97273	(12, 20)	68514	(13, 45)	96726	(15, 35)	87330
(6, 18)	94893	(7, 37)	99605	(9, 15)	68847	(10, 37)	97564	(12, 21)	72182	(13, 46)	96998	(15, 36)	88360
(6, 19)	95822	(7, 38)	99662	(9, 16)	73607	(10, 38)	97822	(12, 22)	75348	(13, 47)	97246	(15, 37)	89297
(6, 20)	96573	(7, 39)	99711	(9, 17)	77523	(10, 39)	98052	(12, 23)	78095	(13, 48)	97473	(15, 38)	90149
(6, 21)	97182	(7, 40)	99753	(9, 18)	80771	(10, 40)	98257	(12, 24)	80489	(13, 49)	97680	(15, 39)	90926
(6, 22)	97678	(7, 41)	99788	(9, 19)	83485	(10, 41)	98439	(12, 25)	82582	(13, 50)	97869	(15, 40)	91636
(6, 23)	98084	(7, 42)	99820	(9, 20)	85765	(10, 42)	98602	(12, 26)	84419	$n = 14$		(15, 41)	92285
(6, 24)	98417	(7, 43)	99845	(9, 21)	87693	(10, 43)	98747	(12, 27)	86036	(14, 15)	13333	(15, 42)	92878
(6, 25)	98690	(7, 44)	99867	(9, 22)	89331	(10, 44)	98877	(12, 28)	87465	(14, 16)	24419	(15, 43)	93422
(6, 26)	98915	(7, 45)	99886	(9, 23)	90727	(10, 45)	98993	(12, 29)	88730	(14, 17)	33725	(15, 44)	93921
(6, 27)	99101	(7, 46)	99902	(9, 24)	91924	(10, 46)	99096	(12, 30)	89853	(14, 18)	41607	(15, 45)	94379
(6, 28)	99254	(7, 47)	99916	(9, 25)	92952	(10, 47)	99189	(12, 31)	90852	(14, 19)	48331	(15, 46)	94799
(6, 29)	99381	(7, 48)	99928	(9, 26)	93838	(10, 48)	99272	(12, 32)	91743	(14, 20)	54107	(15, 47)	95186
(6, 30)	99485	(7, 49)	99939	(9, 27)	94605	(10, 49)	99347	(12, 33)	92539	(14, 21)	59098	(15, 48)	95542
(6, 31)	99572	(7, 50)	99947	(9, 28)	95269	(10, 50)	99413	(12, 34)	93251	(14, 22)	63434	(15, 49)	95870
(6, 32)	99644	$n = 8$		(9, 29)	95846	$n = 11$		(12, 35)	93890	(14, 23)	67220	(15, 50)	96172
(6, 33)	99704	(8, 9)	22222	(9, 30)	96348	(11, 12)	16667	(12, 36)	94463	(14, 24)	70539	$n = 16$	
(6, 34)	99754	(8, 10)	38400	(9, 31)	96787	(11, 13)	29864	(12, 37)	94978	(14, 25)	73462	(16, 18)	21769
(6, 35)	99795	(8, 11)	50505	(9, 32)	97170	(11, 14)	40476	(12, 38)	95442	(14, 26)	76044	(16, 19)	30341
(6, 36)	99830	(8, 12)	59763	(9, 33)	97505	(11, 15)	49120	(12, 39)	95860	(14, 27)	78333	(16, 20)	37735
(6, 37)	99858	(8, 13)	66972	(9, 34)	97799	(11, 16)	56240	(12, 40)	96238	(14, 28)	80369	(16, 21)	44153
(6, 38)	99882	(8, 14)	72670	(9, 35)	98057	(11, 17)	62162	(12, 41)	96579	(14, 29)	82185	(16, 22)	49754
(6, 39)	99902	(8, 15)	77229	(9, 36)	98283	(11, 18)	67127	(12, 42)	96887	(14, 30)	83809	(16, 23)	54665
(6, 40)	99918	(8, 16)	80916	(9, 37)	98482	(11, 19)	71323	(12, 43)	97166	(14, 31)	85264	(16, 24)	58991
(6, 41)	99932	(8, 17)	83925	(9, 38)	98658	(11, 20)	74890	(12, 44)	97419	(14, 32)	86571	(16, 25)	62818
(6, 42)	99943	(8, 18)	86400	(9, 39)	98812	(11, 21)	77942	(12, 45)	97648	(14, 33)	87747	(16, 26)	66215
(6, 43)	99953	(8, 19)	88451	(9, 40)	98949	(11, 22)	80565	(12, 46)	97856	(14, 34)	88808	(16, 27)	69242
(6, 44)	99961	(8, 20)	90159	(9, 41)	99069	(11, 23)	82831	(12, 47)	98045	(14, 35)	89767	(16, 28)	71946
(6, 45)	99967	(8, 21)	91590	(9, 42)	99175	(11, 24)	84797	(12, 48)	98217	(14, 36)	90635	(16, 29)	74370
(6, 46)	99973	(8, 22)	92795	(9, 43)	99269	(11, 25)	86508	(12, 49)	98373	(14, 37)	91421	(16, 30)	76508
(6, 47)	99977	(8, 23)	93813	(9, 44)	99352	(11, 26)	88003	(12, 50)	98514	(14, 38)	92135	(16, 31)	78510
(6, 48)	99981	(8, 24)	94676	(9, 45)	99426	(11, 27)	89314	$n = 13$		(14, 39)	92783	(16, 32)	80281
(6, 49)	99984	(8, 25)	95411	(9, 46)	99491	(11, 28)	90466	(13, 14)	14286	(14, 40)	93374	(16, 33)	81884
(6, 50)	99987	(8, 26)	96037	(9, 47)	99548	(11, 29)	91481	(13, 15)	26000	(14, 41)	93912	(16, 34)	83337
$n = 7$		(8, 27)	96574	(9, 48)	99599	(11, 30)	92377	(13, 16)	35714	(14, 42)	94403	(16, 35)	84657
(7, 8)	25000	(8, 28)	97034	(9, 49)	99644	(11, 31)	93171	(13, 17)	43849	(14, 43)	95851	(16, 36)	85859
(7, 9)	42424	(8, 29)	97429	(9, 50)	99684	(11, 32)	93875	(13, 18)	50719	(14, 44)	95260	(16, 37)	86954
(7, 10)	55000	(8, 30)	97770	$n = 10$		(11, 33)	94501	(13, 19)	56565	(14, 45)	95635	(16, 38)	87953
(7, 11)	64326	(8, 31)	98063	(10, 11)	18182	(11, 34)	95058	(13, 20)	61573	(14, 46)	95978	(16, 39)	88867
(7, 12)	71392	(8, 32)	98317	(10, 12)	32258	(11, 35)	95555	(13, 21)	65888	(14, 47)	96293	(16, 40)	89703
(7, 13)	76841	(8, 33)	98536	(10, 13)	43357	(11, 36)	95999	(13, 22)	69627	(14, 48)	96581	(16, 41)	90469
(7, 14)	81104	(8, 34)	98726	(10, 14)	52242	(11, 37)	96395	(13, 23)	72881	(14, 49)	96846	(16, 42)	91173
(7, 15)	84480	(8, 35)	98890	(10, 15)	59447	(11, 38)	96750	(13, 24)	75726	(14, 50)	97089	(16, 43)	91819
(7, 16)	87181	(8, 36)	99033	(10, 16)	65354	(11, 39)	97068	(13, 25)	78224	$n = 15$		(16, 44)	92413
(7, 17)	89362	(8, 37)	99157	(10, 17)	70243	(11, 40)	97354	(13, 26)	80424	(15, 16)	12500	(16, 45)	92960
(7, 18)	91135	(8, 38)	99265	(10, 18)	74324	(11, 41)	97610	(13, 27)	82369	(15, 17)	23018	(16, 46)	93464
(7, 19)	92586	(8, 39)	99359	(10, 19)	77755	(11, 42)	97841	(13, 28)	84093	(15, 18)	31944	(16, 47)	93929
(7, 20)	93780	(8, 40)	99441	(10, 20)	80657	(11, 43)	98048	(13, 29)	85625	(15, 19)	39578	(16, 48)	94358
(7, 21)	94769	(8, 41)	99512	(10, 21)	83127	(11, 44)	98235	(13, 30)	86991	(15, 20)	46150	(16, 49)	94754
(7, 22)	95589	(8, 42)	99574	(10, 22)	85239	(11, 45)	98403	(13, 31)	88211	(15, 21)	51843	(16, 50)	95130
(7, 23)	96274	(8, 43)	99628	(10, 23)	87054	(11, 46)	98555	(13, 32)	89304	(15, 22)	56801	$n = 17$	
(7, 24)	96846	(8, 44)	99675	(10, 24)	88619	(11, 47)	98691	(13, 33)	90283	(15, 23)	61139	(17, 18)	11111
(7, 25)	97327	(8, 45)	99716	(10, 25)	89975	(11, 48)	98815	(13, 34)	91164	(15, 24)	64953	(17, 19)	20648
(7, 26)	97731	(8, 46)	99752	(10, 26)	91152	(11, 49)	98926	(13, 35)	91957	(15, 25)	68319	(17, 20)	28889
(7, 27)	98072	(8, 47)	99783	(10, 27)	92178	(11, 50)	99027	(13, 36)	92672	(15, 26)	71301	(17, 21)	36054
(7, 28)	98360	(8, 48)	99810	(10, 28)	93074	$n = 12$		(13, 37)	93318	(15, 27)	73951	(17, 22)	42317
(7, 29)	98603	(8, 49)	99834	(10, 29)	93859	(12, 13)	15385	(13, 38)	93902	(15, 28)	76314	(17, 23)	47620
(7, 30)	98810	(8, 50)	99855	(10, 30)	94548	(12, 14)	27799	(13, 39)	94430	(15, 29)	78428	(17, 24)	52676
(7, 31)	98985	$n = 9$		(10, 31)	95154	(12, 15)	37949	(13, 40)	94910	(15, 30)	80322	(17, 25)	56979
(7, 32)	99134	(9, 10)	20000	(10, 32)	95688	(12, 16)	46340	(13, 41)	95345	(15, 31)	82025	(17, 26)	60807
(7, 33)	99260	(9, 11)	35065	(10, 33)	96159	(12, 17)	53344	(13, 42)	95740	(15, 32)	83559	(17, 27)	64222

TABLE—*Continued*

(n, t)	C(n, t)	(n, t)	C(n, t)	(n, t)	C(n, t)	(n, t)	C(n, t)	(n, t)	C(n, t)	(n, t)	C(n, t)	(n, t)	C(n, t)
(17, 28)	67281	(19, 25)	44360	(21, 27)	41355	(23, 33)	54244	(25, 43)	70434	(28, 34)	33396	(31, 34)	17279
(17, 29)	70027	(19, 26)	49087	(21, 28)	45910	(23, 34)	57325	(25, 44)	72104	(28, 35)	37471	(31, 35)	22157
(17, 30)	72499	(19, 27)	53319	(21, 29)	50080	(23, 35)	60153	(25, 45)	73662	(28, 36)	41233	(31, 36)	26667
(17, 31)	74731	(19, 28)	57120	(21, 30)	53829	(23, 36)	62753	(25, 46)	75117	(28, 37)	44711	(31, 37)	30842
(17, 32)	76750	(19, 29)	60543	(21, 31)	57232	(23, 37)	65147	(25, 47)	76478	(28, 38)	47931	(31, 38)	34715
(17, 33)	78581	(19, 30)	63636	(21, 32)	60330	(23, 38)	67356	(25, 48)	77751	(28, 39)	50919	(31, 39)	38313
(17, 34)	80245	(19, 31)	66436	(21, 33)	63155	(23, 39)	69397	(25, 49)	78943	(28, 40)	53694	(31, 40)	41662
(17, 35)	81759	(19, 32)	68978	(21, 34)	65736	(23, 40)	71285	(25, 50)	80061	(28, 41)	56275	(31, 41)	44782
(17, 36)	83140	(19, 33)	71291	(21, 35)	68101	(23, 41)	73035	n = 26		(28, 42)	58680	(31, 42)	47693
(17, 37)	84401	(19, 34)	73399	(21, 36)	70270	(23, 42)	74650	(26, 27)	07407	(28, 43)	60923	(31, 43)	50413
(17, 38)	85555	(19, 35)	75325	(21, 37)	72261	(23, 43)	76168	(26, 28)	14105	(28, 44)	63018	(31, 44)	52958
(17, 39)	86612	(19, 36)	77086	(21, 38)	74091	(23, 44)	77571	(26, 29)	20179	(28, 45)	64976	(31, 45)	55341
(17, 40)	87582	(19, 37)	78701	(21, 39)	75791	(23, 45)	78877	(26, 30)	25703	(28, 46)	66810	(31, 46)	57576
(17, 41)	88472	(19, 38)	80184	(21, 40)	77353	(23, 46)	80095	(26, 31)	30740	(28, 47)	68527	(31, 47)	59673
(17, 42)	89292	(19, 39)	81547	(21, 41)	78798	(23, 47)	81231	(26, 32)	35343	(28, 48)	70138	(31, 48)	61643
(17, 43)	90046	(19, 40)	82802	(21, 42)	80135	(23, 48)	82293	(26, 33)	39561	(28, 49)	71651	(31, 49)	63497
(17, 44)	90742	(19, 41)	83960	(21, 43)	81375	(23, 49)	83285	(26, 34)	43434	(28, 50)	73072	(31, 50)	65241
(17, 45)	91383	(19, 42)	85028	(21, 44)	82525	(23, 50)	84213	(26, 35)	46996	n = 29		n = 32	
(17, 46)	91976	(19, 43)	86015	(21, 45)	83593	n = 24		(26, 36)	50279	(29, 30)	06667	(32, 33)	06061
(17, 47)	92524	(19, 44)	86929	(21, 46)	84587	(24, 25)	08000	(26, 37)	53311	(29, 31)	12757	(32, 34)	11643
(17, 48)	93031	(19, 45)	87775	(21, 47)	85512	(24, 26)	15174	(26, 38)	56114	(29, 32)	18333	(32, 35)	16797
(17, 49)	93501	(19, 46)	88560	(21, 48)	86374	(24, 27)	21630	(26, 39)	58711	(29, 33)	23452	(32, 36)	21562
(17, 50)	93936	(19, 47)	89288	(21, 49)	87178	(24, 28)	27458	(26, 40)	61121	(29, 34)	28160	(32, 37)	25977
n = 18		(19, 48)	89965	(21, 50)	87929	(24, 29)	32736	(26, 41)	63359	(29, 35)	32491	(32, 38)	30075
(18, 17)	11765	(19, 49)	90594	n = 22		(24, 30)	37529	(26, 42)	65441	(29, 36)	36506	(32, 39)	33883
(18, 18)	21769	(19, 50)	91179	(22, 23)	08696	(24, 31)	41893	(26, 43)	67381	(29, 37)	40213	(32, 40)	37429
(18, 19)	30341	n = 20		(22, 24)	16418	(24, 32)	45875	(26, 44)	69189	(29, 38)	43647	(32, 41)	40734
(18, 20)	37735	(20, 21)	09524	(22, 25)	23304	(24, 33)	49519	(26, 45)	70878	(29, 39)	46835	(32, 42)	43820
(18, 21)	44153	(20, 22)	17884	(22, 26)	29469	(24, 34)	52858	(26, 46)	72457	(29, 40)	49798	(32, 43)	46704
(18, 22)	49754	(20, 23)	25259	(22, 27)	35007	(24, 35)	55926	(26, 47)	73933	(29, 41)	52556	(32, 44)	49404
(18, 23)	54665	(20, 24)	31795	(22, 28)	39998	(24, 36)	58748	(26, 48)	75316	(29, 42)	55127	(32, 45)	51938
(18, 24)	58991	(20, 25)	37611	(22, 29)	44509	(24, 37)	61350	(26, 49)	76613	(29, 43)	57526	(32, 46)	54306
(18, 25)	62818	(20, 26)	42806	(22, 30)	48598	(24, 38)	63753	(26, 50)	78298	(29, 44)	59768	(32, 47)	56534
(18, 27)	58910	(20, 27)	47463	(22, 31)	52315	(24, 39)	65975	n = 27		(29, 45)	61866	(32, 48)	58629
(18, 28)	62333	(20, 28)	51651	(22, 32)	55701	(24, 40)	68034	(27, 28)	07143	(29, 46)	63830	(32, 49)	60600
(18, 29)	65413	(20, 29)	55428	(22, 33)	58792	(24, 41)	69943	(27, 29)	13625	(29, 47)	65672	(32, 50)	62457
(18, 30)	68191	(20, 30)	58845	(22, 34)	61621	(24, 42)	71716	(27, 30)	19524	(29, 48)	67401	n = 33	
(18, 31)	70703	(20, 31)	61943	(22, 35)	64214	(24, 43)	73364	(27, 31)	24906	(29, 49)	69025	(33, 34)	05882
(18, 32)	72979	(20, 32)	64759	(22, 36)	66595	(24, 44)	74899	(27, 32)	29829	(29, 50)	70551	(33, 35)	11314
(18, 33)	75047	(20, 33)	67324	(22, 37)	68787	(24, 45)	76330	(27, 33)	34342	n = 30		(33, 36)	16340
(18, 34)	76929	(20, 34)	69665	(22, 38)	70807	(24, 46)	77665	(27, 34)	38488	(30, 31)	06452	(33, 37)	20998
(18, 35)	78645	(20, 35)	71807	(22, 39)	72671	(24, 47)	78912	(27, 35)	42305	(30, 32)	12363	(33, 38)	25323
(18, 36)	80213	(20, 36)	73770	(22, 40)	74394	(24, 48)	80077	(27, 36)	45826	(30, 33)	17791	(33, 39)	29344
(18, 37)	81647	(20, 37)	75571	(22, 41)	75989	(24, 49)	81168	(27, 37)	49078	(30, 34)	22786	(33, 40)	33090
(18, 38)	82962	(20, 38)	77227	(22, 42)	77467	(24, 50)	82190	(27, 38)	52088	(30, 35)	27393	(33, 41)	36583
(18, 39)	84170	(20, 39)	78752	(22, 43)	78839	n = 25		(27, 39)	54879	(30, 36)	31649	(33, 42)	39846
(18, 40)	85279	(20, 40)	80158	(22, 44)	80114	(25, 26)	07692	(27, 40)	57469	(30, 37)	35588	(33, 43)	42897
(18, 41)	86300	(20, 41)	81457	(22, 45)	81300	(25, 27)	14620	(27, 41)	59878	(30, 38)	39240	(33, 44)	45754
(18, 42)	87241	(20, 42)	82657	(22, 46)	82404	(25, 28)	20879	(27, 42)	62120	(30, 39)	42632	(33, 45)	48432
(18, 43)	88110	(20, 43)	83768	(22, 47)	83433	(25, 29)	26552	(27, 43)	64209	(30, 40)	45786	(33, 46)	50946
(18, 44)	88912	(20, 44)	84798	(22, 48)	84393	(25, 30)	31707	(27, 44)	66160	(30, 41)	48724	(33, 47)	53307
(18, 45)	89653	(20, 45)	85754	(22, 49)	85289	(25, 31)	36404	(27, 45)	67982	(30, 42)	51463	(33, 48)	55528
(18, 46)	90340	(20, 46)	86641	(22, 50)	86127	(25, 32)	40694	(27, 46)	69686	(30, 43)	54022	(33, 49)	57619
(18, 47)	90976	(20, 47)	87466	n = 23		(25, 33)	44622	(27, 47)	71282	(30, 44)	56414	(33, 50)	59589
(18, 48)	91565	(20, 48)	88233	(23, 24)	08333	(25, 34)	48225	(27, 48)	72778	(30, 45)	58653	n = 34	
(18, 49)	92113	(20, 49)	88948	(23, 25)	15771	(25, 35)	51538	(27, 49)	74181	(30, 46)	60751	(34, 35)	05714
(18, 50)	92621	(20, 50)	89614	(23, 26)	22436	(25, 36)	54589	(27, 50)	75499	(30, 47)	62720	(34, 36)	11003
n = 19		n = 21		(23, 27)	28428	(25, 37)	57403	n = 28		(30, 48)	64568	(34, 37)	15907
(19, 20)	10000	(21, 22)	09091	(23, 28)	33834	(25, 38)	60004	(28, 29)	06897	(30, 49)	66306	(34, 38)	20462
(19, 21)	18719	(21, 23)	17120	(23, 29)	38725	(25, 39)	62412	(28, 30)	13176	(30, 50)	67940	(34, 39)	24700
(19, 22)	26364	(21, 24)	24242	(23, 30)	43162	(25, 40)	64644	(28, 31)	18910	n = 31		(34, 40)	28648
(19, 23)	33099	(21, 25)	30588	(23, 31)	47199	(25, 41)	66716	(28, 32)	24157	(31, 32)	06250	(34, 41)	32332
(19, 24)	39061	(21, 26)	36263	(23, 32)	50880	(25, 42)	68641	(28, 33)	28971	(31, 33)	11992	(34, 42)	35774

TABLE—*Concluded*

(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$	(n, t)	$C(n, t)$
(34, 43)	38095	(35, 48)	49075	(37, 40)	14737	(38, 38)	38800	(40, 46)	25076	(42, 48)	24074	(45, 48)	12319
(34, 44)	42011	(35, 49)	51411	(37, 41)	19008	(38, 49)	41517	(40, 47)	28422	(42, 49)	27320	(45, 49)	15977
(34, 45)	44840	(35, 50)	53613	(37, 42)	23002	(38, 50)	44081	(40, 48)	31578	(42, 50)	30387	(45, 50)	19437
(34, 46)	47497	$n = 36$		(37, 43)	26744	$n = 39$		(40, 49)	34557	$n = 43$		$n = 46$	
(34, 47)	49993	(36, 37)	05405	(37, 44)	30253	(39, 40)	05000	(40, 50)	37369	(43, 44)	04545	(46, 47)	04255
(34, 48)	52343	(36, 38)	10430	(37, 45)	33547	(39, 41)	09674	$n = 41$		(43, 45)	08821	(46, 48)	08273
(34, 49)	54555	(36, 39)	15107	(37, 46)	36643	(39, 42)	14048	(41, 42)	04762	(43, 46)	12846	(46, 49)	12071
(34, 50)	56640	(36, 40)	19469	(37, 47)	39557	(39, 43)	18147	(41, 43)	09227	(43, 47)	16640	(46, 50)	15664
$n = 35$		(36, 41)	23542	(37, 48)	42302	(39, 44)	21994	(41, 44)	13420	(43, 48)	20221	$n = 47$	
(35, 36)	05556	(36, 42)	27350	(37, 49)	44889	(39, 45)	25608	(41, 45)	17361	(43, 49)	23602	(47, 48)	04167
(35, 37)	10709	(36, 43)	30916	(37, 50)	47332	(39, 46)	29007	(41, 46)	21070	(43, 50)	26800	(47, 49)	08106
(35, 38)	15497	(36, 44)	34258	$n = 38$		(39, 47)	32208	(41, 47)	24565	$n = 44$		(47, 50)	11833
(35, 39)	19953	(36, 45)	37395	(38, 39)	05128	(39, 48)	35225	(41, 48)	27860	(44, 45)	04444	$n = 48$	
(35, 40)	24107	(36, 46)	40343	(38, 40)	09913	(39, 49)	38071	(41, 49)	30971	(44, 46)	08630	(48, 49)	04082
(35, 41)	27984	(36, 47)	43116	(38, 41)	14384	(39, 50)	40760	(41, 50)	33910	(44, 47)	12577	(48, 50)	07945
(35, 42)	31608	(36, 48)	45727	(38, 42)	18567	$n = 40$		$n = 42$		(44, 48)	16302	$n = 49$	
(35, 43)	35000	(36, 49)	48188	(38, 43)	22487	(40, 41)	04878	(42, 43)	04651	(44, 49)	19821	(49, 50)	04000
(35, 44)	38179	(36, 50)	50510	(38, 44)	26164	(40, 42)	09145	(42, 44)	09019	(44, 50)	23149		
(35, 45)	41161	$n = 37$		(38, 45)	29617	(40, 43)	13727	(42, 45)	13127	$n = 45$			
(35, 46)	43962	(37, 38)	05263	(38, 46)	32864	(40, 44)	17745	(42, 46)	16993	(45, 46)	04348		
(35, 47)	46596	(37, 39)	10165	(38, 47)	35920	(40, 45)	21522	(42, 47)	20637	(45, 47)	08448		

REFERENCES

[1] A. C. Cohen, Jr., "Estimation of the Poisson parameter in truncated samples," *J. Amer. Stat. Assn., Vol. 49 (1954), pp. 158–168*

[2] F. N David and N. L. Johnson, "The truncated Poisson," *Biometrics*, Vol. 8 (1952), pp. 275–285.

[3] Charles Jordan, *Calculus of Finite Differences*, Chelsea, New York, 1950.

[4] E. L. Lehmann, *Notes on the Theory of Estimation*, University of California Press, Berkeley, 1950.

[5] P. G. Moore, "The estimation of the Poisson parameter from a truncated distribution," *Biometrika*, Vol. 39 (1952), pp. 247–251.

[6] P. G. Moore, "A note on truncated Poisson distributions," *Biometrics*, Vol. 10 (1954), pp. 402–406.

[7] R. L. Plackett, "The truncated Poisson distribution," *Biometrics*, Vol. 9 (1953) pp. 485–488.

[8] P. R. Rider, "Truncated Poisson distributions," *J. Amer Stat Assn* , Vol. 48 (1953), pp. 826–830.

[9] L. H. C. Tippett, "A modified method of counting particles," *Proc. Roy. Soc London Ser. A*, Vol. 137 (1932), pp. 434–446.

[10] J W. Tukey, "Sufficiency, truncation, and selection," *Ann. Math. Stat.*, Vol. 20, (1949), pp. 309–310.

ON BALANCING IN FACTORIAL EXPERIMENTS[1]

BY B. V. SHAH

University of Bombay

1. Introduction and Summary. R. C. Bose [1] has considered the problem of balancing in symmetrical factorial experiments. In all the designs considered in that paper, the block size is a power of S, the number of levels of a factor. The purpose of the present paper is to consider a general class of designs, where a 'complete balance' is achieved over different effects and interactions. It is proved in this paper (Theorems 4.1 and 4.2) that if a 'complete balance' is achieved over each order of interaction, the design must be a partially balanced incomplete block design. Its parameters are found. The usual method of analysis (of a PBIB design [2]) which is not so simple, can be simplified a little for these designs (section 5), on account of the balancing of the interactions of various orders. The simplified method of analysis is illustrated by a worked out example 5.1. Finally, the problem of balancing is dealt with for asymmetrical factorial experiments also. Incidentally, it may be observed that the generalised quasifactorial designs discussed by C. R. Rao [4] are the same as found by the author, from considerations of balancing.

2. Some lemmas regarding C-matrix and orthogonal contrasts. Let there be v treatments replicated $r_1, r_2, \cdots, r_v$ times respectively, in b blocks of k plots each. Let n_{ij} be the number of times the ith treatment occurs in the jth block; ($i = 1, 2, \cdots, v$; $j = 1, 2, \cdots, b$). Then $\mathbf{N} = [n_{ij}]$ is the incidence matrix of the design. It is assumed that every n_{ij} is either zero or one. The set up assumed is that the yield of a plot in the jth block having the ith treatment is $\mu + \alpha_i + t_j + \epsilon_{ij}$ where μ is the over-all effect, α_i is the effect of the ith block, t_j is the effect of the jth treatment and ϵ_{ij} is the experimental error. ϵ_{ij}'s are assumed to be independent normal variates with zero mean and variance σ^2. Let Q_i be the adjusted treatment yield (adjusted for block effects) of the ith treatment, and $\hat{t}_i$ be a solution for t_i of the least square equations. Let $\mathbf{Q}$, $\mathbf{t}$ and $\hat{\mathbf{t}}$ denote the column vectors $(Q_1, Q_2, \cdots, Q_v)$, $(t_1, t_2, \cdots, t_v)$, and $(\hat{t}_1, \hat{t}_2, \cdots, \hat{t}_v)$ respectively.

It is well known that

$$\mathbf{Q} = \mathbf{C}\hat{\mathbf{t}} \tag{2.1}$$

and the variance-covariance matrix of Q is

$$\sigma^2\mathbf{C}. \tag{2.2}$$

where

$$\mathbf{C} = \text{diag}\,(r_1, r_2, \cdots, r_v) - \frac{1}{k}\mathbf{NN}', \tag{2.3}$$

Received October 17, 1957.

[1] This work was supported by a Research Training Scholarship from the Government of India.

diag $(r_1, r_2, \cdots, r_v)$ stands for a diagonal matrix, with diagonal elements $r_1, r_2, \cdots, r_v$.

If $l'l = 1$, the contrast $l't$ will be called a normalised contrast.

LEMMA 2.1. *Let* $l_1't$, $l_2't$, $\cdots$, $l_{v-1}'t$ *be* $v - 1$ *estimable normalised orthogonal contrasts* (l_i's *are v-vectors*), *such that*

$$V(l_i'\hat{t}) = \sigma^2/\theta_i \tag{2.4}$$

$$\text{Cov}\,(l_i'\hat{t}, l_j'\hat{t}) = 0 \qquad i \neq j \tag{2.5}$$

then (i) the **C**-matrix defined in (2.3) is given by

$$C = \sum_{q=1}^{v-1} \theta_q l_q l_q'. \tag{2.6}$$

(ii) Estimate of $l_i't$ is given by

$$l_i'\hat{t} = l_i'Q/\theta_i. \tag{2.7}$$

PROOF. Let E_{mn} denote an $m \times n$ matrix, all the elements of which are unity and

$$\left[l_1 \mid l_2 \mid \cdots \left| l_{v-1} \right| \frac{1}{\sqrt{v}} E_{v1}\right] = \left[L_1 \left| \frac{1}{\sqrt{v}} E_{v1}\right.\right] = L, \tag{2.8}$$

then

$$LL' = I_v = L'L, \tag{2.9}$$

where I_v denotes a $v \times v$ identity matrix. From (2.1) and (2.9) we have

$$\begin{aligned} Q &= CLL'\hat{t}. \\ L'Q &= L'CL(L'\hat{t}), \end{aligned} \tag{2.10}$$

but

$$E_{1v}Q = O \quad \text{and} \quad E_{1v}C = O; \tag{2.11}$$

hence (2.10) reduces to

$$L_1'Q = L_1'CL_1(L_1'\hat{t}). \tag{2.12}$$

From (2.2) it follows that the variance-covariance matrix of $L_1'Q$ is

$$L_1'CL_1\,\sigma^2. \tag{2.13}$$

By hypothesis each of $l_1't \cdots l_{v-1}'t$ is estimable, therefore $(L_1'CL_1)$ must have rank $v - 1$. Hence its inverse exists.

$$(L_1'\hat{t}) = (L_1'CL_1)^{-1}L'Q \tag{2.14}$$

and

$$V(L_1'\hat{t}) = (L_1'CL_1)^{-1}\sigma^2. \tag{2.15}$$

Comparing with (2.4) we have

$$(L_1'\,CL_1)^{-1} = \text{diag}\left(\frac{1}{\theta_1}, \frac{1}{\theta_2}, \cdots, \frac{1}{\theta_{v-1}}\right) \tag{2.16}$$

$$\mathbf{L}_1' \mathbf{C} \mathbf{L}_1 = \text{diag}(\theta_1, \theta_2, \cdots, \theta_{v-1}). \tag{2.17}$$

(2.11) and (2.17) imply that $\theta_1, \theta_2, \cdots, \theta_{v-1}' 0$ are canonical roots of $\mathbf{C}$, and $\mathbf{l}_1, \mathbf{l}_2, \cdots, \mathbf{l}_{v-1}, (1/\sqrt{v}) \mathbf{E}_{v1}$ are corresponding canonical vectors. Hence $\mathbf{C}$ is given by

$$\mathbf{C} = \sum_{q=1}^{v-1} \theta_q \mathbf{l}_q \mathbf{l}_q'. \tag{2.18}$$

Also from (2.14) and (2.16) it follows

$$\mathbf{L}_1' \hat{\mathbf{t}} = \text{diag}\left(\frac{1}{\theta_1}, \frac{1}{\theta_2}, \cdots, \frac{1}{\theta_{v-1}}\right) \mathbf{L}_1' \mathbf{Q}. \tag{2.19}$$

This proves (2.7).

LEMMA 2.2. *In case some of the* θ's *in Lemma* 2.1 *are equal say* $\theta_1 = \theta_2 = \cdots = \theta_r = \theta$, *then there will be infinitely many sets of normalised orthogonal vectors corresponding to the canonical root* θ. *The variance-covariance matrix of contrasts corresponding to any such set will be*

$$\frac{\sigma^2}{\theta} \mathbf{I}_r$$

and representation of $\mathbf{C}$ *as given by Lemma* (2.1) *is unique;* i.e. if $\mathbf{l}_1, \cdots, \mathbf{l}_r$; *and* $\mathbf{n}_1, \cdots, \mathbf{n}_r$ *are any two sets, then*

$$\sum_{i=1}^{r} \mathbf{l}_i \mathbf{l}_i' = \sum_{i=1}^{r} \mathbf{n}_i \mathbf{n}_i'.$$

The proof follows easily from observing that

$$[\mathbf{n}_1 \mid \mathbf{n}_2 \mid \cdots \mid \mathbf{n}_r] = [\mathbf{l}_1 \mid \mathbf{l}_2 \mid \cdots \mid \mathbf{l}_r] \cdot \mathbf{A}, \tag{2.20}$$

where $\mathbf{A}$ is an $r \times r$ orthogonal matrix.

3. Definition of 'complete balance'. In a factorial experiment with m factors $F_1, F_2, \cdots, F_m$ each at S levels, if the treatments are denoted by $(x_1 x_2, \cdots, x_m)$ where x_i is the level of ith factor ($x_i = 0, 1, 2, \cdots, S - 1$); then a contrast $\sum C_{x_1\, x_2, \cdots, x_m} (x_1 x_2, \cdots, x_m)$ (Summation is over all $x_1 x_2, \cdots, x_m$) belongs to $(q - 1)$th order interaction between the factors $F_{j_1}, F_{j_2}, \cdots, F_{j_q}$, if $C_{x_1, x_2, \cdots, x_m}$ depends only on $x_{j_1}, x_{j_2}, \cdots, x_{j_q}$ and $\sum C_{x_1 x_2, \cdots, x_m}$, summed over the levels of any one of these q factors, is zero.

Bose [1] has defined balance over a particular order of interaction in symmetric factorial experiments. In general, that definition is not interpretable, e.g. when a number of levels S is not a power of a prime, or the block size is not a power of S. So a more general definition is necessary.

DEFINITION 3.1. We shall define that a 'complete balance' is achieved over a set of n normalised orthogonal contrasts $\mathbf{l}_n^1 \mathbf{t}, \cdots, \mathbf{l}_n^1 \mathbf{t}$ if and only if the variance-covariance matrix of their estimates is

$$\frac{\sigma^2}{\theta} \mathbf{I}_n .$$

Definition 3.2. A more obvious definition of 'complete balance' over a set of vectors or contrasts represented by them is that every linear combination of these vectors giving a normalised contrast is estimated with the same variance say σ^2/θ.

Theorem 3.1. *Two Definitions* 3.1 *and* 3.2 *are equivalent.*

We will now say that complete balance is achieved over $(q - 1)$th order of interaction; if a complete set of $\binom{m}{q}(S - 1)^q$ normalised orthogonal contrasts has variance-covariance matrix (σ^2/θ_q) I, or if every normalised contrast belonging to the q factor interaction is estimated with the same variance σ^2/θ_q.

4. Balanced factorial designs and PBIB. Let there be m factors each at S levels in a symmetric factorial experiment. Let $\mathbf{L}_q$ be $S^m \times \binom{m}{q}(S - 1)^q$ matrix formed by a complete set of $\binom{m}{q}(S - 1)^q$ normalised orthogonal vectors forming q factor interactions with the variance of the estimate of any normalised contrast belonging to a q factor interaction equal to σ^2/θ_q; $q = 1, 2, \cdots, m$. Further let us assume that the covariance between the estimates of any two contrasts belonging to the ith and the jth $(i \neq j)$ orders of interactions is zero.

From Lemmas 1.1 and 1.2 C is uniquely represented and given by

$$\mathbf{C} = \sum_{q=1}^{m} \theta_q \mathbf{L}_q \mathbf{L}_q', \tag{4.1}$$

which can also be written as

$$\mathbf{C} = \left[\sum_{q=1}^{m} \theta_q f_{ij}^q\right], \qquad i, j = 1, 2, \cdots, S^m, \tag{4.2}$$

where f_{ij}^q is the element of $\mathbf{L}_q\mathbf{L}_q'$ corresponding to ith row and jth column.

Let the ith and jth treatments be $(x_1, x_2, \cdots, x_m)$ and $(y_1, y_2, \cdots, y_m)$ respectively, and let

$$(0, 0, \cdots, 0) \text{ and } (\underbrace{0, 0, \cdots, 0}_{p \text{ times}}, \underbrace{1, 1, \cdots, 1}_{(m-p) \text{ times}})$$

be the rth and sth treatments respectively. In the ith and jth treatments suppose exactly p factors occur at the same level. Say $x_{i_1} = y_{i_1}$, $x_{i_2} = y_{i_2}$, $\cdots$, $x_{i_p} = y_{i_p}$, and rest of the x_i's are not equal to the corresponding y_i's. Now interchange the levels $x_1, x_2, \cdots, x_m$ with zeros, i.e., in any treatment if the ith factor occurs at level x_i replace it by zero and if it occurs at level zero replace it by x_i. Perform this change for all the treatments. So naturally $y_{i_1}, y_{i_2}, \cdots, y_{i_p}$ will be changed to zeros. Now in the same manner as x_i's, interchange the remaining levels y_i's with ones. After these interchanges call the i_1th factor as the first factor, i_2th factor as the second factor, $\cdots$, and lastly i_pth factor as

the pth factor and the other $(m - p)$ factors as $(p + 1)$th to mth factors; and rewrite all the treatments accordingly. Then it is obvious that the ith treatment becomes $(0\ 0, \cdots, 0)$ and the jth treatment,

$$(\underbrace{0\ 0, \cdots, 0}_{p \text{ times}}, \underbrace{1\ 1, \cdots, 1}_{(m-p) \text{ times}}).$$

It is obvious that interchanges of levels or renaming the levels of any factor does not alter the order of an interaction; so also the permutation or renaming of factors. Hence the above changes will not alter the order of any interaction.

After renaming the treatments arrange them in the original order. This will mean permutation of rows of $\mathbf{L}_q$. Let the rearranged matrix be $\mathbf{M}_q$. Then the rth row of $\mathbf{M}_q$ is the ith row of $\mathbf{L}_q$ and the sth row of $\mathbf{M}_q$ is the jth row of $\mathbf{L}_q$. Let $\mathbf{L}_q\mathbf{L}_q' = [l_{ij}]$ and $\mathbf{M}_q\mathbf{M}_q' = [m_{ij}]$ $i, j = 1, 2, \cdots, s_m$. Then it is evident that

$$l_{ij} = m_{rs}. \tag{4.3}$$

It is easy to see that M_q also gives a complete set of normalised orthogonal contrasts belonging to the $(q - 1)$th order or q-factor interactions. Hence from Lemma 2.2

$$\mathbf{L}_q\mathbf{L}_q' = \mathbf{M}_q\mathbf{M}_q' \quad \text{i.e.} \quad l_{rs} = m_{rs}. \tag{4.4}$$

Hence

$$l_{ij} = l_{rs}. \tag{4.5}$$

This shows that f^q_{ij} depends only on the exact number of factors say p, which occur at the same level in both ith and jth treatments. Let us denote it by f^q_p, $p = 0, 1, \cdots, m$; $p = m$ denotes all levels equal $(i = j)$ and f^q_m is a diagonal element.

Equating the two forms of $\mathbf{C}$ (2.3) and (4.2) with $v = S^m$, we obtain

$$\text{Diag}\,(r_1, r_2, \cdots, r_v) - \frac{1}{k}\mathbf{NN}' = \left[\sum_{q=1}^{m} \theta_q f^q_{ij}\right], \tag{4.6}$$

Equating the elements we get

$$\sum_{q=1}^{m} \theta_q f^q_{ii} = r_i\left(1 - \frac{1}{k}\right) \tag{4.7}$$

and

$$\sum_{q=1}^{m} \theta_q f^q_{ij} = -\frac{\lambda_{ij}}{k}\ (i = j) \tag{4.8}$$

where λ_{ij} equals number of times ith and jth treatment occur together.

Using (4.5), (4.7) and (4.8) we have

$$r_1 = r_2 = , \cdots, r_v = \frac{k}{k-1}\sum_{q=1}^{m} \theta_q f^q_m = r \quad \text{say}, \tag{4.9}$$

and if ith and jth treatments have p factors at the same level,

$$(4.10)\qquad -\frac{\lambda_{ij}}{k} = \sum_{q=1}^{m} \theta_q f_p^q = -\frac{\lambda_p}{k} \quad \text{say.}$$

Now (4.9) and (4.10) imply that the design must be a partially balanced incomplete block design. The definition of P.B.I.B. was first given by Bose and Nair [2] and later generalised by Nair and Rao [3].

Parameters b, k, r, being selected to satisfy combinatorial properties of the design and $v = S^m$, pth associates of any treatment will be all the treatments which have exactly p factors at the same level as in the given treatment. Hence

$$(4.11)\qquad n_p = \binom{m}{p}(S-1)^{m-p} \qquad p = 0, 1, \cdots, m-1$$

and

$$(4.12)\quad p_{ij}^k = \sum_u \binom{k}{u}\binom{m-k}{i-u}\binom{m-k-i+u}{j-u}(S-1)^{k-u}(S-2)^{(m-k-i-j+2u)},$$

where summation extends over all the values of u which are less than or equal to minimum of k, i, j and for which $m + 2u > k + i + j$. Parameters $\lambda_0, \lambda_1, \cdots, \lambda_{m-1}$ are given by

$$(4.13)\qquad \begin{bmatrix} f_0^0 & f_0^1 & \cdots & f_0^m \\ f_1^0 & f_1^1 & \cdots & f_1^m \\ \vdots & \vdots & & \vdots \\ f_m^0 & f_m^1 & \cdots & f_m^m \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix} = -\frac{1}{k}\begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}$$

where $\lambda_m = -r(k-1)$

$$f_p^0 = \frac{1}{S^m} \quad \text{for} \quad p = 0, 1, \cdots, m.$$

and θ_0 is a dummy parameter always equal to zero, introduced to simplify the inverse relation. (4.13) can be shortly written as

$$\mathbf{F}(m)\cdot\boldsymbol{\theta}(m) = -\frac{1}{k}\,\boldsymbol{\lambda}(m)$$

As it will be shown later in section 7 the inverse relation of (4.13) exists and can be written as

$$(4.14)\qquad \boldsymbol{\theta}(m) = -\frac{1}{k}\,[\mathbf{F}(m)]^{-1}\boldsymbol{\lambda}(m).$$

Therefore it also follows that in every P.B.I.B. with parameters as given above 'complete balance' over each order of interaction is achieved.

Hence we have the following theorems.

THEOREM 4.1. *Every P.B.I.B. design with parameters as given in* (4.11) *and* (4.12) *achieves a 'complete balance' over each order of interaction.*

the pth factor and the other $(m - p)$ factors as $(p + 1)$th to mth factors; and rewrite all the treatments accordingly. Then it is obvious that the ith treatment becomes $(0\ 0, \cdots, 0)$ and the jth treatment,

$$(\underbrace{0\ 0, \cdots, 0}_{p \text{ times}}, \underbrace{1\ 1, \cdots, 1}_{(m-p) \text{ times}}).$$

It is obvious that interchanges of levels or renaming the levels of any factor does not alter the order of an interaction; so also the permutation or renaming of factors. Hence the above changes will not alter the order of any interaction.

After renaming the treatments arrange them in the original order. This will mean permutation of rows of $\mathbf{L}_q$. Let the rearranged matrix be $\mathbf{M}_q$. Then the rth row of $\mathbf{M}_q$ is the ith row of $\mathbf{L}_q$ and the sth row of $\mathbf{M}_q$ is the jth row of $\mathbf{L}_q$. Let $\mathbf{L}_q\mathbf{L}_q' = [l_{ij}]$ and $\mathbf{M}_q\mathbf{M}_q' = [m_{ij}]$ $i, j = 1, 2, \cdots, s_m$. Then it is evident that

$$l_{ij} = m_{rs}. \tag{4.3}$$

It is easy to see that M_q also gives a complete set of normalised orthogonal contrasts belonging to the $(q - 1)$th order or q-factor interactions. Hence from Lemma 2.2

$$\mathbf{L}_q\mathbf{L}_q' = \mathbf{M}_q\mathbf{M}_q' \quad \text{i.e.} \quad l_{rs} = m_{rs}. \tag{4.4}$$

Hence

$$l_{ij} = l_{rs}. \tag{4.5}$$

This shows that f_{ij}^q depends only on the exact number of factors say p, which occur at the same level in both ith and jth treatments. Let us denote it by f_p^q, $p = 0, 1, \cdots, m$; $p = m$ denotes all levels equal $(i = j)$ and f_m^q is a diagonal element.

Equating the two forms of $\mathbf{C}$ (2.3) and (4.2) with $v = S^m$, we obtain

$$\text{Diag}(r_1, r_2, \cdots, r_v) - \frac{1}{k}\mathbf{NN}' = \left[\sum_{q=1}^{m} \theta_q f_{ij}^q\right], \tag{4.6}$$

Equating the elements we get

$$\sum_{q=1}^{m} \theta_q f_{ii}^q = r_i\left(1 - \frac{1}{k}\right) \tag{4.7}$$

and

$$\sum_{q=1}^{m} \theta_q f_{ij}^q = -\frac{\lambda_{ij}}{k}\ (i = j) \tag{4.8}$$

where λ_{ij} equals number of times ith and jth treatment occur together.

Using (4.5), (4.7) and (4.8) we have

$$r_1 = r_2 = , \cdots, r_v = \frac{k}{k-1}\sum_{q=1}^{m} \theta_q f_m^q = r \quad \text{say}, \tag{4.9}$$

and if ith and jth treatments have p factors at the same level,

$$-\frac{\lambda_{ij}}{k} = \sum_{q=1}^{m} \theta_q f_p^q = -\frac{\lambda_p}{k} \quad \text{say.} \tag{4.10}$$

Now (4.9) and (4.10) imply that the design must be a partially balanced incomplete block design. The definition of P.B.I.B. was first given by Bose and Nair [2] and later generalised by Nair and Rao [3].

Parameters b, k, r, being selected to satisfy combinatorial properties of the design and $v = S^m$, pth associates of any treatment will be all the treatments which have exactly p factors at the same level as in the given treatment. Hence

$$n_p = \binom{m}{p} (S-1)^{m-p} \qquad p = 0, 1, \cdots, m-1 \tag{4.11}$$

and

$$p_{ij}^k = \sum_u \binom{k}{u}\binom{m-k}{i-u}\binom{m-k-i+u}{j-u} (S-1)^{k-u}(S-2)^{(m-k-i-j+2u)}, \tag{4.12}$$

where summation extends over all the values of u which are less than or equal to minimum of k, i, j and for which $m + 2u > k + i + j$. Parameters $\lambda_0, \lambda_1, \cdots, \lambda_{m-1}$ are given by

$$\begin{bmatrix} f_0^0 & f_0^1 & \cdots & f_0^m \\ f_1^0 & f_1^1 & \cdots & f_1^m \\ \vdots & \vdots & & \vdots \\ f_m^0 & f_m^1 & \cdots & f_m^m \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix} = -\frac{1}{k} \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix} \tag{4.13}$$

where $\lambda_m = -r(k-1)$

$$f_p^0 = \frac{1}{S^m} \quad \text{for} \quad p = 0, 1, \cdots, m.$$

and θ_0 is a dummy parameter always equal to zero, introduced to simplify the inverse relation. (4.13) can be shortly written as

$$\mathbf{F}(m) \cdot \boldsymbol{\theta}(m) = -\frac{1}{k} \boldsymbol{\lambda}(m).$$

As it will be shown later in section 7 the inverse relation of (4.13) exists and can be written as

$$\boldsymbol{\theta}(m) = -\frac{1}{k} [\mathbf{F}(m)]^{-1} \boldsymbol{\lambda}(m). \tag{4.14}$$

Therefore it also follows that in every P.B.I.B. with parameters as given above 'complete balance' over each order of interaction is achieved.

Hence we have the following theorems.

THEOREM 4.1. *Every* P.B.I.B. *design with parameters as given in* (4.11) *and* (4.12) *achieves a 'complete balance' over each order of interaction.*

THEOREM 4.2. *If in a design*

(i) *'complete balance' is obtained over each order of interaction*

(ii) *covariance between the estimates of any two contrasts belonging to different orders of interactions is zero; and*

(iii) *the number of plots is the same in every block; then the design must be a* P.B.I.B. *with parameters given above.*

COROLLARY 4.2.1. *In any design with S treatments if complete balance is achieved over all contrasts then the* C-*matrix is of the form given by*

$$C = \theta\left(I_S - \frac{1}{S} E_{SS}\right) \tag{4.15}$$

COROLLARY 4.2.2. *In any design if complete balance is achieved over all contrasts and if the block size is the same for all the blocks, then the design must be a balanced incomplete block design.*

From (4.15) it follows that if $m = 1$,

$$f_0^1 = -\frac{1}{S}; \qquad f_1^1 = \frac{S-1}{S} \tag{4.16}$$

and hence

$$F(1) = \frac{1}{S}\begin{bmatrix} 1 & -1 \\ 1 & S-1 \end{bmatrix}. \tag{4.17}$$

5. Analysis. Let us consider a symmetrical factorial design which is a P.B.I.B. of the type defined in section 4. Then as in (4.1)

$$C = \sum_{q=1}^{m} \theta_q L_q L_q' \tag{5.1}$$

where θ's are given by (4.14) as

$$\theta(m) = -\frac{1}{k}[F(m)]^{-1} \cdot \lambda(m). \tag{5.2}$$

Hence if $l't$ is any normalised contrast belonging to $(q-1)$th order interaction, applying Lemma 1.1 we have

$$l'\hat{t} = l'Q/\theta_q \tag{5.3}$$

$$V(l'\hat{t}) = \sigma^2/\theta_q \tag{5.4}$$

and

$$\text{S.S. due to } l't = \frac{(l'Q)^2}{\theta_q}. \tag{5.5}$$

Now if T_i is the yield of the ith treatment, and t is a column vector $(T_1, T_2, \cdots, T_v)$ and we suppose that the experiment is a randomised block design with r replications, then

$$l'\hat{t} = l't/r \tag{5.6}$$

$$V(\mathbf{l}'\hat{\mathbf{t}}) = \sigma^2/r \tag{5.7}$$

and

$$\text{S.S. due to } \mathbf{l}'\hat{\mathbf{t}} = \frac{(\mathbf{l}'\mathbf{T})^2}{r}. \tag{5.8}$$

Hence by comparing (5.3), (5.4) and (5.5) with (5.6), (5 7) and (5.8) respectively, we obtain the following procedure for analysis:

(i) calculation of Q

(ii) calculation of sums of squares for each order of interaction separately, as if it were a randomised block experiment but using Q in place of T

(iii) calculation of θ_q's by using (5 2)

(iv) correcting S.S. obtained in (ii) by θ_q's instead of by r.

If we have a quasifactorial experiment or if it is necessary for some purpose, we will require estimates of individual treatment effects and variances of elementary treatment comparisons. For that we know by (2.19),

$$\mathbf{L}_q'\hat{\mathbf{t}} = \frac{1}{\theta_q}\mathbf{L}_q'\mathbf{Q} \tag{5.9}$$

Hence

$$\sum_{q=1}^{m}\mathbf{L}_q\mathbf{L}_q'\hat{\mathbf{t}} = \left[\sum_{q=1}^{m}\frac{1}{\theta_q}\mathbf{L}_q\mathbf{L}_q'\right]\mathbf{Q}. \tag{5.10}$$

Since

$$\left(\mathbf{L}_1 \mid \mathbf{L}_2 \mid \cdots \mid \mathbf{L}_m \mid \frac{1}{\sqrt{S^m}}\mathbf{E}_{S^m 1}\right)$$

is an orthogonal matrix, (5.10) simplifies to

$$\left[\mathbf{I}_v - \frac{1}{v}\mathbf{E}_{vv}\right]\hat{\mathbf{t}} = \left[\sum_{q=1}^{m}\frac{1}{\theta_q}\mathbf{L}_q\mathbf{L}_q'\right]\mathbf{Q}. \tag{5.11}$$

where $v = s^m$. Put $\mathbf{E}_{1v}\hat{\mathbf{t}} = \mathbf{0}$ and we obtain a solution given by

$$\hat{\mathbf{t}} = \left[\sum_{q=1}^{m}\frac{1}{\theta_q}\mathbf{L}_q\mathbf{L}_q'\right]\mathbf{Q} \tag{5.12}$$

$$\hat{\mathbf{t}} = \mathbf{M}\mathbf{Q} \quad \text{say.}$$

Let U_s be defined as follows

$$\mathbf{F}(m)\begin{bmatrix} 0 \\ 1/\theta_1 \\ 1/\theta_2 \\ \vdots \\ 1/\theta_m \end{bmatrix} = \begin{bmatrix} U_0 \\ U_1 \\ U_2 \\ \vdots \\ U_m \end{bmatrix}. \tag{5.13}$$

Then as in (4.5) $U_0, U_1, \cdots, U_m$ are the elements of M. The element in the

ith row and jth column is U_p if the ith and jth treatments have exactly p factors at the same level. Hence (5.12) simplifies to

$$\hat{t}_j = U_m Q_j + \sum_{i=1}^{m} U_i S_i(Q_j) \tag{5.14}$$

where $S_i(Q_j)$ is sum of Q_j's corresponding to the treatments which are ith associates of t_j as defined in (4.11). From solutions (5.14) it is easy to see that, if t_i and t_j are pth associates

$$V(\hat{t}_i - \hat{t}_j) = 2\sigma^2(U_m - U_p). \tag{5.15}$$

EXAMPLE 5.1. Consider example with two factors A and B each at three levels

$$V = 3^2 \quad b = 6 \quad K = 6 \quad r = 4$$

$$n_0 = n_1 = 4 \quad \lambda_0 = 3 \quad \lambda_1 = 2$$

Block No.	Treatments					
1	(1 0)	(2 0)	(0 1)	(2 1)	(0 2)	(1 2)
2	(0 0)	(1 0)	(1 1)	(2 1)	(0 2)	(2 2)
3	(0 0)	(2 0)	(0 1)	(1 1)	(1 2)	(2 2)
4	(1 0)	(2 0)	(0 1)	(1 1)	(0 2)	(2 2)
5	(0 0)	(2 0)	(1 1)	(2 1)	(1 2)	(1 2)
6	(0 0)	(0 1)	(1 0)	(2 1)	(1 2)	(2 2)

Using the formulas in section 7.

$$\mathbf{F}(2) = \frac{1}{9}\begin{bmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 4 & 4 \end{bmatrix}$$

$$[\mathbf{F}(2)]^{-1} = \begin{bmatrix} 4 & 4 & 1 \\ -2 & 1 & 1 \\ 1 & -2 & 1 \end{bmatrix}$$

Apply (5.2)

$$\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = -\frac{1}{6}\begin{bmatrix} 4 & 4 & 1 \\ -2 & 1 & 1 \\ 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ -20 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 7/2 \end{bmatrix}$$

Let Q_{ij} denote adjusted treatment yield of (ij) and

$$Q_{.j} = \sum_{i=0}^{2} Q_{ij}$$

$$Q_{i.} = \sum_{j=0}^{2} Q_{ij}.$$

Then

$$\text{Main effect of } A = \sum_{i=0}^{2} Q_{i.}^2/4.3.$$

$$\text{Main effect of } B = \sum_{j=0}^{2} Q_{.j}^2/4.3.$$

$$\text{Interaction } AB = \frac{2}{7}\left(\sum Q_{ij}^2 - \frac{\sum Q_{.j}^2}{3} - \frac{\sum Q_{i.}^2}{3}\right).$$

Also

$$\mathbf{F}(2)\begin{bmatrix} 0 \\ 1/\theta_1 \\ 1/\theta_2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 4 & 4 \end{bmatrix}\begin{bmatrix} 0 \\ 1/4 \\ 2/7 \end{bmatrix} = \begin{bmatrix} -1/42 \\ -1/28 \\ 5/21 \end{bmatrix}.$$

Hence using (5.14)

$$\hat{t}_j = \tfrac{5}{21}Q_j - \tfrac{1}{42}S_0(Q_j) - \tfrac{1}{28}S_1(Q_j)$$

and using (5.17) we get

$$V(\hat{t}_i - \hat{t}_j) = \tfrac{3}{14}\sigma^2 \quad \text{if } t_i \text{ and } t_j \text{ are 0th associates;}$$
$$= \tfrac{17}{42}\sigma^2 \quad \textit{otherwise.}$$

6. $S_1^{m_1}S_2^{m_2}, \cdots, S_h^{m_h}$ Factorial experiment. Some matrix operators are defined *to derive certain further results.*

Operator '$\times$' denotes the Kronecker product of matrices defined by

$$(6.1) \qquad \mathbf{A} \times \mathbf{B} = [a_{ij}] \times \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B}, \cdots, a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B}, \cdots, a_{2n}\mathbf{B} \\ \vdots & \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B}, \cdots, a_{mn}\mathbf{B} \end{bmatrix}.$$

The operator '$\otimes$' denotes the symbolic kroneker product of suffixes defined by the following illustrations.

$$(6.2) \qquad \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} \otimes \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} \lambda_{00} \\ \lambda_{01} \\ \lambda_{10} \\ \lambda_{11} \end{bmatrix}$$

and

$$(6.3) \qquad \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \otimes \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \theta_{00} \\ \theta_{01} \\ \theta_{02} \\ \theta_{10} \\ \theta_{11} \\ \theta_{12} \end{bmatrix}.$$

THEOREM 6.1. *If in a* $S_1^{m_1}S_2^{m_2}, \cdots, S_h^{m_h}$ *factorial experiment*

(i) *any contrast belonging to the interaction involving* q_i *factors at* S_i *levels* ($i = 1, 2, \cdots, h$) *is estimated with the same variance say* $\sigma^2/\theta_{q_1 q_2, \cdots, q_h}$

(ii) *the estimates of all effects and interaction are all uncorrelated and*

(iii) *the block size is a constant equal to* k *say; then the design must be a* PBIB *with relevant parameters and conversely.*

If any two treatments have exactly p_i factors (each at S_i level) at the same level for $i = 1, 2, \cdots, h$; they will be called $(p_1 p_2, \cdots, p_h)$th associates. Then we have

$$n_{p1, p2, \cdots, ph} = \prod_{i=1}^{h} \binom{m_i}{P_i} (S_i - 1)^{m_i - P_i} \tag{6.4}$$

and the relations between θ's and λ's are

$$\begin{aligned} &F(m_1) \times F(m_2) \times, \cdots, \times F(m_h) \cdot \theta(m_1) \otimes \theta(m_2) \otimes \cdots \otimes \theta(m_h) \\ &\quad = -\frac{1}{k} \lambda(m_1) \otimes \lambda(m_2) \otimes \cdots \otimes \lambda(m_h), \end{aligned} \tag{6.5}$$

$$\begin{aligned} &\theta(m_1) \otimes \theta(m_2) \otimes \cdots \otimes \theta(m_h) \\ &\quad = -\frac{1}{k} [F(m_1)]^{-1} \times [F(m_2)]^{-1} \times \cdots \times [F(m_h)]^{-1} \\ &\qquad \cdot \lambda(m_1) \otimes \lambda(m_2) \otimes \cdots \otimes \lambda(m_h) \end{aligned} \tag{6.6}$$

where $\theta_{00, \cdots, 0} = 0$ and $\lambda_{m_1 m_2, \cdots, m_h} = -r(k - 1)$.

PROOF. The theorem can be proved for $h = 2$ exactly on the same lines as section 4 and relation (6.5) can be obtained by noting that the matrix representing an interaction of $(q_1 + q_2)$ factors out of $m_1 + m_2$ factors can be expressed as the Kronecker product of two matrices representing interactions of q_1 and q_2 factors, out of m_1 and m_2 factors respectively; and then using properties of the Kronecker product of matrices. And the result can be easily generalised for any value of h. (6.5) and (6.6) can be used to simplify the analysis of many asymmetrical factorial experiments. For example the design of plan 6.9 of Cochran and Cox [11] has parameters $v = 3.2^2$, $b = 6$, $r = 3$, $k = 6$ and $\lambda_{00} = 1$, $\lambda_{10} = 3$, $\lambda_{01} = 2$, $\lambda_{11} = 0$, $\lambda_{02} = 1$, $\lambda_{12} = -15$; hence θ's can be calculated as $\theta_{11} = \theta_{01} = \theta_{10} = 3$ and $\theta_{02} = 8/3$, $\theta_{12} = 5/3$ and the analysis can be performed as in section 5.

7. Evaluation of F(m) and $[F(m)]^{-1}$. Put $m_1 = m_2 = \cdots = m_h = 1$ in (6.7) and write $F(m_i)$ as $F_i(1)$ to avoid ambiguity. Then (6.7) becomes

$$\begin{aligned} &F_1(1) \times F_2(1) \times \cdots \times F_h(1) \cdot \theta(1) \otimes \theta(1) \otimes \cdots \otimes \theta(1) \\ &\quad = -\frac{1}{k} \lambda(1) \otimes \lambda(1) \otimes \cdots \otimes \lambda(1). \end{aligned} \tag{7.1}$$

From (4.17) we have

$$F_i(1) = \frac{1}{S_i} \begin{bmatrix} 1 & -1 \\ 1 & S_i - 1 \end{bmatrix}. \tag{7.2}$$

Hence

(7.3) $$[\mathbf{F}_i(1)]^{-1} = \begin{bmatrix} S_i - 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Hence (7.1) and its inverse relation can be written as

(7.4) $$\lambda_{d_1 d_2, \ldots, d_h} = \frac{-k}{\prod_{i=1}^{h} S_i} \sum \prod_{i=1}^{h} G_i(c_i d_i) \theta_{c_1 c_2, \ldots, c_h}$$

and

(7.5) $$\theta_{d_1 d_2, \ldots, d_h} = -\frac{1}{k} \sum \prod_{i=1}^{h} H_i(c_i d_i) \lambda_{c_1 c_2, \ldots, c_h},$$

where c_i and d_i take values 0 or 1; the summation is over all the values of $(c_1 c_2, \cdots, c_h)$ and

$$G_i(11) = S_i - 1 = H_i(0, 0)$$
$$G_i(10) = \quad -1 = H_i(0, 1)$$
$$G_i(00) = G_i(01) = \quad 1 = H_i(10) = F_i(11).$$

Now put $S_1 = S_2 = \cdots = S_h = S$ in (7.4) and $\theta_{c_1 c_2, \ldots, c_h} = \theta_q$ where $q =$ number of ones in $(c_1 c_2, \cdots, c_h)$; on simplifying the coefficient of θ_q on the right side of (7.4) is given by

(7.6) $$\sum' \prod_{i=1}^{h} G_i(c_i d_i)$$

where $\sum'$ is summation for those values of $(c_1 c_2, \cdots, c_h)$ which have exactly q ones and $h - q$ zeros. Now if the number of ones in $(d_1 d_2, \cdots, d_h)$ is p, then it is easy to prove that,

(7.7) $$\sum' \prod_{i=1}^{h} G_i(c_i d_i) = \sum_i^* \binom{p}{i} \binom{h-p}{q-i} (-1)^{q-i} (S-1)^i$$

where $\sum_i^*$ is summation over all the values of i such that

$$\max(0, p + q - h) \leqq i \leqq \min(p, q).$$

Hence if there is balance over each order of interaction, $\lambda_{d_1 d_2, \ldots, d_h}$ depends only on the exact number of factors (say p) which occur at the same level. This must be so, as it was proved in section 4. Now writing $\lambda_{d_1 d_2, \ldots, d_h}$ as λ_p (7.4) becomes

(7.8) $$\lambda_p = \frac{-k}{S^h} \sum_{q=0}^{m} \sum_i^* \binom{p}{i} \binom{m-p}{q-i} (-1)^{q-i} (S-1)^i \theta_q.$$

Comparing (7.8) and (4.13) with $m = h$ we obtain

(7.9) $$f_p^q = \frac{1}{S^m} \sum_i^* \binom{p}{i} \binom{m-p}{q-i} (-1)^{q-i} (S-1)^i.$$

Working similiarly with (7.5) we obtain

$$(7.10)\quad \theta_q = -\frac{1}{K}\sum_{p=0}^{m}\sum_j{}^*\binom{m-q}{j}\binom{q}{m-p-j}(-1)^{m-p-j}(S-1)^j\lambda_p,$$

where $\sum_j^*$ is summation over all the values of j such that

$$\max\ (0,\ m-p-q) \leq j \leq \min\ (m-p, m-q).$$

Hence the inverse relation of (4.13) exists and is given by (7.10). If g_p^q is an element in the $(p+1)$th row and $(q+1)$th column of $[F(m)]^{-1}$ then on comparing (7.10) and (4.14), we have

$$(7.11)\qquad g_p^q = \sum_j{}^*\binom{m-p}{j}\binom{p}{m-q-j}(-1)^{m-q-j}(S-1)^j.$$

Equations (7.9) and (7.11) are not convenient for writing down the matrices $F(m)$ and $[F(m)]^{-1}$. But the following relations, easily derivable from them will enable us to write out these matrices easily, along with a check.

$$(7.12)\qquad g_0^q = \binom{m}{q}(S-1)^{m-q}$$

$$(7.13)\qquad g_p^0 = (-1)^p\ (S-1)^{m-p}$$

$$(7.14)\qquad g_p^m = 1$$

$$(7.15)\qquad g_m^q = \binom{m}{q}(-1)^{m-q}$$

$$(7.16)\qquad g_{p-1}^{q-1} = g_p^{q-1} + g_{p-1}^q + (S-1)g_p^q$$

$$(7.17)\qquad g_p^q = S^m \cdot f_{m-p}^{m-q}.$$

8. Remarks. It should be noted that a general class of quasifactorial designs as defined by C. R. Rao [4] has the same parameters as given in (7.4). Hence the variance of a treatment contrast for any design belonging to that class can beo btained from (7.5).

Two factor designs in the above class form an important group. Their analysis can be done by using (7.4) and (7.5) with $h = 2$ and the method given in section 5. It will yield the same expressions as given by C. R. Rao and K. R. Nair in [10]. They are, therefore, not reproduced here.

Secondly construction of PBIB designs with parameters as required in the above designs is considered by M. N. Vartak [5] D. A. Sprott [6] and C. R. Rao [4].

Furthermore in the above design if $\lambda_{00} = \lambda_{01}$ or λ_{10} then $\theta_{11} = \theta_{01}$ or θ_{10} and the design becomes a group divisible PBIB.

All the designs mentioned in this paper can be successfully used by introducing Pseudo-factors. The method of introducing Pseudo-factors is discussed by Kramer and Bradley [12] for factorial experiments in group divisible PBIB.

9. Acknowledgment. The author is very grateful to Professor M. C. Chakrabarti for suggesting the problem and for his guidance.

REFERENCES

[1] R. C Bose, "Mathematics of symmetric factorial experiments", *Sankhya* Vol. 8 (1947), pp. 107–166.

[2] R C Bose and K. R. Nair, "Partially balanced designs," *Sankhya* Vol. 4 (1939), pp. 337-372

[3] K. R. Nair and C. R. Rao, "A note on partially balanced incomplete block design", *Science and Culture* Vol. 7 (1941–42) pp. 568–569.

[4] C. R. Rao, "General class of quasi-factorial designs," *Sankhya* Vol. 17 (1956) pp. 165–174.

[5] M. N. Vartak, "On an application of kronecker product of matrices to statistical designs," *Ann. Math. Stat.* 26 (1955), pp. 420–438.

[6] D. A. Sprott, "Some series of partially balanced incomplete block designs," *Canadian J. Math.* Vol. 7 (1955), pp. 369–381.

[7] K. R. Nair and C. R. Rao, "Confounded designs for asymmetrical factorial experiments", *Science and Culture* Vol. 7 (1941–42) pp 313–314.

[8]—"Confounded designs for the $k \times p^m \times q^n \times \cdots$ type of factorial experiment," *Science and Culture* Vol. 7 (1941–42), pp. 361–362.

[9]—"A general class of quasi-factorial designs leading to confounded factorial experiments," *Science and Culture* Vol. 7 (1941–42), pp. 457–458.

[10]—"Confounding in asymmetrical factorial experiments," J.R.S S. (Series B) Vol. 10 (1948) pp. 109–131.

[11] W. G. Cochran and G. R. Cox, Experimental Designs. John Wiley and Sons, Inc , New York. 1950.

[12] C. Y. Kramer and R. A. Bradley, "Intra block analysis for factorials in two associate class group divisible designs," *Ann. Math Stat.* Vol. 28 (1957), pp. 349–361.

A TABLE FOR COMPUTING TRIVARIATE NORMAL PROBABILITIES

By George P. Steck

Sandia Corporation

1. Introduction. For convenience in the following discussion let X, Y, and Z be random variables with a trivariate normal distribution such that $EX = EY = EZ = 0$, $EX^2 = EY^2 = EZ^2 = 1$, $EXY = \rho_{12}$, $EXZ = \rho_{13}$, $EYZ = \rho_{23}$, let $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ denote the probability that $X \leqq h$, $Y \leqq k$, $Z \leqq m$, and let $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ denote the probability that $X \geqq h$, $Y \geqq k$, $Z \geqq m$. Several tables have been prepared from which certain particular values of the trivariate normal integral can be obtained. A tabulation of the area of hyperspherical simplices is given by H. Ruben [1]. The function Ruben has tabulated as $\bar{u}_n(x)$ is, for the case $n = 3$, equal to $C(0, 0, 0; 1/x, 1/x, 1/x)$ and the tabulation is for $x = 2(1)11$. This probability can be computed directly, however, as a special case of the well-known formula (for example, see [2]).

$$(1.1)\quad \begin{aligned} C(0, 0, 0; \rho_{12}, \rho_{13}, \rho_{23}) &= D(0, 0, 0; \rho_{12}, \rho_{13}, \rho_{23}) \\ &= \frac{1}{4\pi}(2\pi - \arccos \rho_{12} - \arccos \rho_{13} - \arccos \rho_{23}) \end{aligned}$$

Short tabulations of $C(h, h, h; 1/2, 1/2, 1/2)$ have been published by D. Teichroew [3] for $h\sqrt{2} = 0(.01)6.09$ and by P. N. Somerville [11] for $h = 0(.1)2(.5)3$. In addition to these published tables, there are some unpublished tables [4] giving $C(h, h, h; \rho, \rho, \rho)$ for $\rho = 1/(1 + \sqrt{3})$ and $\frac{1}{4}$, $h = 0(.1)3(.5)8$ and for $\rho = 0(.1)0.9$, $h = 0(.2)1$.

Methods for computing $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ have been given by M. G. Kendall [5], R. L. Plackett [6], and S. C. Das [7]. The method of Kendall is to express the trivariate normal density as the inverse of its characteristic function obtaining $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ as a six-dimensional integral. The part of the integral involving the ρ_{ij} is expanded in a power series and the result integrated term by term. The resulting series coverges slowly, however, when the ρ_{ij} are large. Plackett's method, on the other hand, is to consider $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ as a function of the ρ_{ij} and write it as a line integral from $(\rho_{12}, \rho_{13}, \rho_{23}^*)$ to $(\rho_{12}, \rho_{13}, \rho_{23})$ where ρ_{23}^* is chosen to give a degenerate trivariate normal density so that $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}^*)$ becomes a bivariate normal integral. The result of this procedure is that $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ can be expressed as a sum of lower dimensional normal integrals and an integral which must be evaluated by numerical integration.

The method of Das reduces the trivariate integral to a single integral which is then evaluated numerically provided the correlations are such that their product is positive and each is numerically greater than the product of the other two.

In this paper $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ is expressed in terms of the univariate

Received November 8, 1957; revised January 24, 1958.

normal integral, the T-function, which is tabulated by D. B. Owen [8, 9], and the function $S(h, a, b)$ which is tabulated here. Although the reduction of $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ is given in terms of the T-function, it is also possible to give it in terms of the V-function tabulated by C. Nicholson [12] and by the National Bureau of Standards [13], or the L-function tabulated by Karl Pearson [14] and by the National Bureau of Standards [13]. The V and L-functions are related to the T-function by the expressions

$$(1.2) \qquad V(h, ah) = \frac{\arctan a}{2\pi} - T(h, a),$$

$$(1.3) \qquad \begin{aligned} L(h, k; \rho) &= \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \int_h^\infty \int_k^\infty \exp\left[-\tfrac{1}{2}(x^2 + y^2 - 2\rho xy)/(1-\rho^2)\right] dx\, dy \\ &= 1 - \tfrac{1}{2}[G(h) + G(k) + \delta_{hk}] - T\left(h, \frac{k - \rho h}{h\sqrt{1-\rho^2}}\right) \\ &\quad - T\left(k, \frac{h - \rho k}{k\sqrt{1-\rho^2}}\right), \end{aligned}$$

where (this is the same δ defined equivalently by (2.3))

$$\delta_{hk} = \begin{cases} \text{if } h < 0 \text{ or } k < 0 \text{ but not both,} \\ \text{otherwise.} \end{cases}$$

For $h > 0$, $a > 0$, $b > 0$, $S(h, a, b) - (1/4\pi)\arctan (b/(1 + a^2 + a^2b^2)^{1/2})$ is the probability that three independent, standardized, normal variables will lie in the region between the planes $x = 0$, $x - bz = 0$, $y = 0$, and $y = h$ and beyond (in the sense that $z \geqq ay$) the plane $z - ay = 0$, i.e., will lie in the truncated infinite wedge shown in Figure 1.

2. Summary of formulas. The fundamental formulas for $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ are:

Case (i): $h \geqq 0, \quad k \geqq 0, \quad m \geqq 0 \quad$ or $\quad h \leqq 0, \quad k \leqq 0, \quad m \leqq 0,$

$$(2.1) \qquad \begin{aligned} C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) &= \tfrac{1}{2}[(1 - \delta_{a_1c_1})G(h) + (1 - \delta_{a_2c_2})G(k) \\ &\quad + (1 - \delta_{a_3c_3})G(m)] - \tfrac{1}{2}[T(h, a_1) + T(h, c_1) + T(k, a_2) + T(k, c_2) \\ &\quad + T(m, a_3) + T(m, c_3)] - [S(h, a_1, b_1) + S(h, c_1, d_1) \\ &\quad + S(k, a_2, b_2) + S(k, c_2, d_2) + S(m, a_3, b_3) + S(m, c_3, d_3)], \end{aligned}$$

Case (ii): $h \geqq 0, \quad k \geqq 0, \quad m < 0 \quad$ or $\quad h \leqq 0, \quad k \leqq 0, \quad m > 0,$

$$(2.2) \qquad \begin{aligned} C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) &= \tfrac{1}{2}[G(h) + G(k) - \delta_{hk}] - T(h, a_1) - T(k, c_2) \\ &\quad - C(h, k, -m; \rho_{12}, -\rho_{13}, -\rho_{23}), \end{aligned}$$

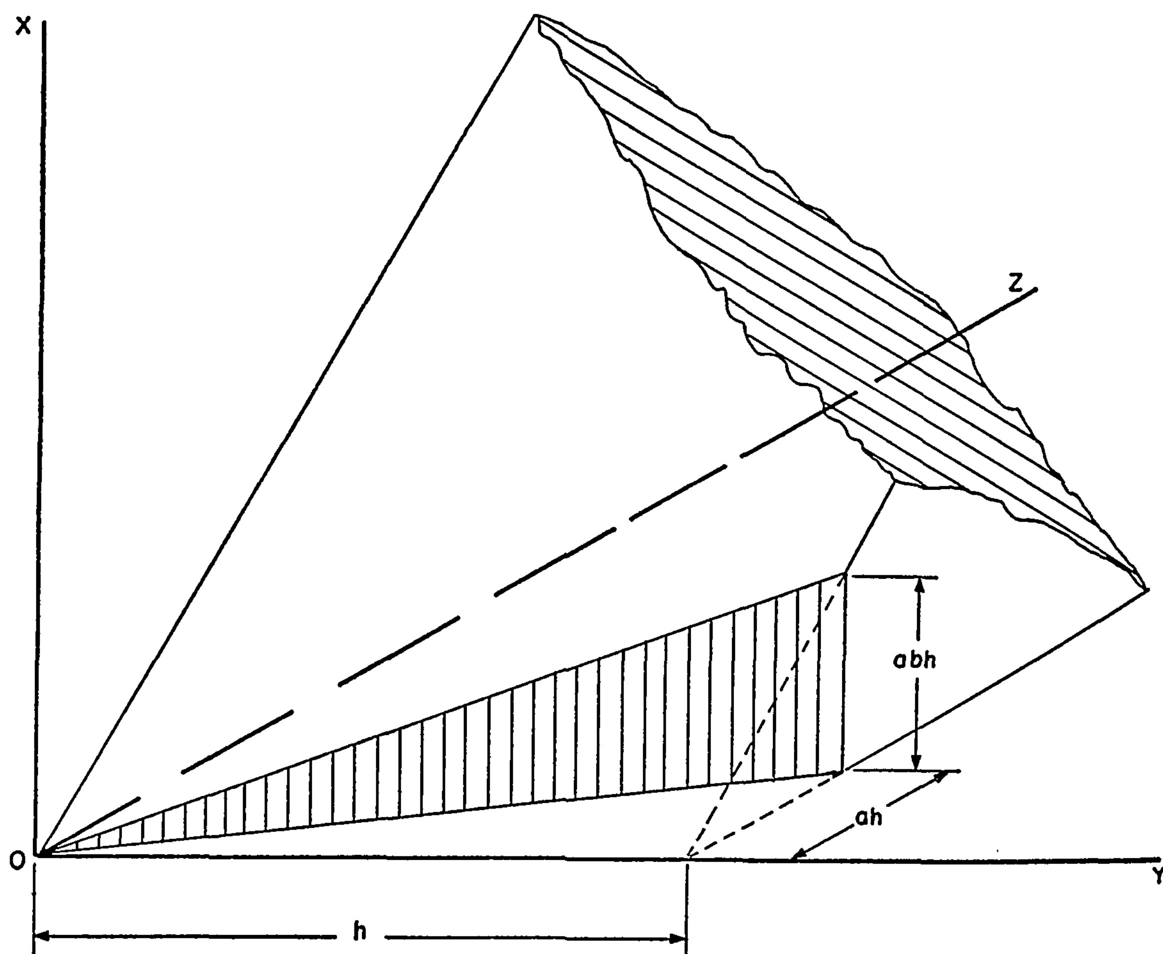


FIG. 1. Volume over which $S(h, a, b) - (1/4\pi)\arctan(b / \sqrt{1 + a^2 + a^2b^2})$ gives the integral of the trivariate normal distribution.

where

$$G(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{x} e^{-x^2/2}\, dx, \; T(h, a) = \frac{1}{2\pi} \int_0^a \frac{e^{-\frac{h^2}{2}(1+x^2)}}{1 + x^2}\, dx,$$

$$a_1 = \frac{k - h\,\rho_{12}}{h(1 - \rho_{12}^2)^{1/2}}, \qquad a_2 = \frac{m - k\rho_{23}}{k(1 - \rho_{23}^2)^{1/2}}, \qquad a_3 = \frac{h - m\rho_{13}}{m(1 - \rho_{13}^2)^{1/2}},$$

$$c_1 = \frac{m - h\rho_{13}}{h(1 - \rho_{13}^2)^{1/2}}, \qquad c_2 = \frac{h - k\rho_{12}}{k(1 - \rho_{12}^2)^{1/2}}, \qquad c_3 = \frac{k - m\rho_{23}}{m(1 - \rho_{23}^2)^{1/2}},$$

$$b_1 = \frac{(1 - \rho_{12}^2)(m - h\rho_{13}) - (\rho_{23} - \rho_{12}\,\rho_{13})(k - h\rho_{12})}{(k - h\rho_{12})\Delta^{1/2}},$$

$$(2.3) \quad d_1 = \frac{(1 - \rho_{13}^2)(k - h\rho_{12}) - (\rho_{23} - \rho_{12}\,\rho_{13})(m - h\rho_{13})}{(m - h\rho_{13})\Delta^{1/2}},$$

$$b_2 = \frac{(1 - \rho_{23}^2)(h - k\rho_{12}) - (\rho_{13} - \rho_{12}\,\rho_{23})(m - k\rho_{23})}{(m - k\rho_{23})\Delta^{1/2}},$$

$$d_2 = \frac{(1 - \rho_{12}^2)(m - k\rho_{23}) - (\rho_{13} - \rho_{12}\,\rho_{23})(h - k\rho_{12})}{(h - k\rho_{12})\Delta^{1/2}},$$

$$b_3 = \frac{(1 - \rho_{13}^2)(k - m\rho_{23}) - (\rho_{12} - \rho_{13}\,\rho_{23})(h - m\rho_{13})}{(h - m\rho_{13})\Delta^{1/2}},$$

$$d_3 = \frac{(1 - \rho_{23}^2)(h - m\rho_{13}) - (\rho_{12} - \rho_{13}\,\rho_{23})(k - m\rho_{23})}{(k - m\rho_{23})\Delta^{1/2}},$$

$$\Delta = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\,\rho_{13}\,\rho_{23},$$

$$\delta_{xy} = \begin{cases} 0 \text{ if } (\operatorname{sgn} x)(\operatorname{sgn} y) = 1 \\ +1 \text{ otherwise} \end{cases},$$

and

$$\operatorname{sgn} x = \begin{cases} 1 & x \geqq 0 \\ -1 & x < 0 \end{cases}.$$

The S-function is tabulated for $0 < b \leqq 1$, but it is possible to obtain values for $1 < b < \infty$ by use of one of the following formulas, $a > 0$, $b > 0$:

$$S(h, a, b) = [G(h) - \tfrac{1}{2}]\, T(ah, b) - [G(hab) - \tfrac{1}{2}]\, T(ah, 1/a) + S(hab, 1/b, 1/a), \tag{2.4}$$

$$S(h, a, b) = (\tfrac{1}{4})G(h) + [G(hab) - \tfrac{1}{2}]\, T(h, a) - S(hab, 1/ab, a) - S(h, ab, 1/b). \tag{2.5}$$

if $a > 1$, $b > 1$ then (2.4) should be used, and if $0 < a \leqq 1$, $b > 1$ then (2.5) should be used. Values for negative h, a, or b may be obtained by using

$$S(-h, a, b) = S(\infty, a, b) - S(h, a, b), \tag{2.6}$$

$$S(h, -a, b) = S(h, a, b), \tag{2.7}$$

$$S(h, a, -b) = -S(h, a, b). \tag{2.8}$$

Note that (2.4) and (2.5) require both a and b to be positive and hence when a or b is negative (2.7) or (2.8) should be applied before (2.4) or (2.5).

Other useful formulas are:

$$\begin{aligned} &S(0, a, b) = \tfrac{1}{2}\, S(\infty, a, b), \qquad S(h, 0, b) = \frac{1}{2\pi}\, G(h) \arctan b, \\ &S(h, a, 0) = 0, \qquad S(\infty, a, b) = \frac{1}{2\pi} \arctan\left[\frac{b}{(1 + a^2 + a^2b^2)^{1/2}}\right], \\ &S(h, \infty, b) = 0, \\ &S(h, a, \infty) = \begin{cases} \tfrac{1}{2}[\tfrac{1}{2}G(h) + T(h, |a|)] - \dfrac{\arctan |a|}{2\pi}, & h \geqq 0 \\ \tfrac{1}{2}[\tfrac{1}{2}G(h) - T(h, |a|)], & h < 0. \end{cases} \end{aligned} \tag{2.9}$$

Equations (2.1) and (2.2) can be easily rewritten in terms of the V-function

by use of (1.2); however, in order to reduce the computation it should be noted that

$$\arctan a_1 + \arctan c_2 = \arctan\left(\sqrt{1-\rho_{12}^2}/\rho_{12}\right).$$

Similar expressions hold for the pairs (a_2, c_3) and (a_3, c_1).

Rewriting equations (2.1) and (2.2) in terms of the L-function gives

Case (i): $h \geqq 0, \quad k \geqq 0, \quad m \geqq 0 \quad$ or $\quad h \leqq 0, \quad k \leqq 0, \quad m \leqq 0,$

$$\begin{aligned} C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) &= (1 - \tfrac{1}{2}\delta_{a_1c_1})G(h) + (1 - \tfrac{1}{2}\delta_{a_2c_2})G(k) \\ &+ (1 - \tfrac{1}{2}\delta_{a_3c_3})G(m) + \tfrac{1}{4}(\delta_{hk} + \delta_{hm} + \delta_{km}) \\ &+ \tfrac{1}{2}[L(h, k; \rho_{12}) + L(h, m; \rho_{13}) + L(k, m; \rho_{23}) - 3] \\ &- [S(h, a_1, b_1) + S(h, c_1, d_1) + S(k, a_2, b_2) + S(k, c_2, d_2) \\ &+ S(m, a_3, b_3) + S(m, c_3, d_3)], \end{aligned} \tag{2.1'}$$

Case (ii): $h \geqq 0, \quad k \geqq 0, \quad m < 0 \quad$ or $\quad h \leqq 0, \quad k \leqq 0, \quad m > 0,$

$$\begin{aligned} C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) &= L(h, k; \rho_{12}) + G(h) + G(k) - 1 \\ &- C(h, k, -m; \rho_{12}, -\rho_{13}, -\rho_{23}). \end{aligned} \tag{2.2'}$$

3. Derivation of the relationship between the trivariate normal integral and the tabulated function. The density function for the standardized trivariate normal distribution is

$$\begin{aligned} f(x, y, z; \rho_{12}, \rho_{13}, \rho_{23}) &= \left(\frac{1}{2\pi}\right)^{3/2} \frac{1}{\Delta^{1/2}} \exp\left[-\tfrac{1}{2}(A_{11}x^2 + A_{22}y^2\right. \\ &\left. + A_{33}z^2 + 2A_{12}xy + 2A_{13}xz + 2A_{23}yz)\right], \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} A_{11} &= \frac{1-\rho_{23}^2}{\Delta}, \quad A_{22} = \frac{1-\rho_{13}^2}{\Delta}, \quad A_{33} = \frac{1-\rho_{12}^2}{\Delta}, \\ A_{12} &= \frac{\rho_{13}\rho_{23}-\rho_{12}}{\Delta}, \quad A_{13} = \frac{\rho_{12}\rho_{23}-\rho_{13}}{\Delta}, \quad A_{23} = \frac{\rho_{12}\rho_{13}-\rho_{23}}{\Delta}, \end{aligned} \tag{3.2}$$

and

$$\Delta = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}.$$

The definition of $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ given earlier is equivalent to

$$C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) = \int_{-\infty}^{h}\int_{-\infty}^{k}\int_{-\infty}^{m} f(x, y, z; \rho_{12}, \rho_{13}, \rho_{23})\, dx\, dy\, dz. \tag{3.3}$$

Let $G(x)$, $T(h, a)$ be as defined in (2.3). It will be convenient to have an alternative form of $T(h, a)$. This is given in [8] and is

$$T(h, a) = \frac{\arctan a}{2\pi} + \tfrac{1}{2}G(h) - \tfrac{1}{4} - \int_0^h G(ax)G'(x)\, dx. \tag{3.4}$$

Also, from [9],

(3.5) $T(h, a) = \frac{1}{2}[G(h) + G(ah)] - G(h)G(ah) - T(ah, 1/a), a > 0.$

Finally, let

$$S(h, a, b) = \int_{-\infty}^{h} T(as, b)G'(s)\, ds. \tag{3.6}$$

It will also be convenient to have an alternative form of (3.6). If the T-function is replaced by its integral representation as given by (2.3) and the order of integrations reversed, the result is

$$S(h, a, b) = \frac{b}{2\pi}\int_0^1 \frac{G[h(1 + a^2 + a^2b^2y^2)^{1/2}]}{(1 + b^2y^2)(1 + a^2 + a^2b^2y^2)^{1/2}}\, dy. \tag{3.7}$$

Integration of (3.6) by parts gives (2.4), and substituting (3.5) into (3.6) and integrating gives (2.5).

The relation between $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ and the S-function can be shown as follows. If h, k, and m are all nonnegative (or nonpositive) and if 0/0 is taken as one, then it can be shown that

$$\begin{aligned} P(X \leqq h, Y \leqq k, Z \leqq m) &= P\left(X \leqq h, Y \leqq \frac{k}{h} X, Z \leqq \frac{m}{h} X\right) \\ &+ P\left(X \leqq \frac{h}{k} Y, Y \leqq k, Z \leqq \frac{m}{k} Y\right) + P\left(X \leqq \frac{h}{m} Z, Y \leqq \frac{k}{m} Z, Z \leqq m\right). \end{aligned} \tag{3.8}$$

Since these three probabilities are all similar in form, it is sufficient to consider only the last. Let the conditional probability, given $Z = s$, that $X \leqq hs/m$ and $Y \leqq ks/m$ be denoted by $A(s)$; then

$$A(s) = B\left(\frac{h - m\rho_{13}}{m(1 - \rho_{13}^2)^{1/2}}\, s,\ \frac{k - m\rho_{23}}{m(1 - \rho_{23}^2)^{1/2}}\, s;\ \frac{\rho_{12} - \rho_{13}\,\rho_{23}}{((1 - \rho_{13}^2)(1 - \rho_{23}^2))^{1/2}}\right). \tag{3.9}$$

where

$$B(h, k; \rho) = \frac{1}{2\pi(1 - \rho^2)^{1/2}} \int_{-\infty}^{h}\int_{-\infty}^{k} \cdot \exp\left[-\tfrac{1}{2}(x^2 - 2\rho xy + y^2)/(1 - \rho^2)\right] dx\, dy. \tag{3.10}$$

Therefore,

$$P\left(X \leqq \frac{h}{m} Z, Y \leqq \frac{k}{m} Z, Z \leqq m\right) = \int_{-\infty}^{m} A(s)G'(s)\, ds. \tag{3.11}$$

However, it is shown in [9] that

$$\begin{aligned} B(h, k; \rho) &= \tfrac{1}{2}[G(h) + G(k)] - T\left(h, \frac{k - \rho h}{h(1 - \rho^2)^{1/2}}\right) \\ &\quad - T\left(k, \frac{h - \rho k}{k(1 - \rho^2)^{1/2}}\right) - \tfrac{1}{2}\delta_{hk} \end{aligned} \tag{3.12}$$

where δ_{hk} has already been defined by (2.3); therefore, expressing (3.9) in the

form of (3.12), substituting in (3.11), and noting (3.4) it follows that

$$
\begin{aligned}
P\left(X \leqq \frac{h}{m} Z, Y \leqq \frac{k}{m} Z, Z \leqq m\right) &= \tfrac{1}{2}(1 - \delta_{a_3 c_3}) G(m) \\
&\quad - \tfrac{1}{2}[T(m, a_3) + T(m, c_3)] - \int_{-\infty}^{m} G'(s) T(a_3 s, b_3)\, ds \\
&\quad - \int_{-\infty}^{m} G'(s) T(c_3 s, d_3)\, ds,
\end{aligned} \tag{3.13}
$$

where a_3, b_3, d_3, and $\delta_{a_3 c_3}$ have already been defined by (2.3). The integrals on the right side of (3.13) are, noting (3.6), $S(m, a_3, b_3)$ and $S(m, c_3, d_3)$. Thus

$$
\begin{aligned}
P\left(X \leqq \frac{h}{m} Z, Y \leqq \frac{k}{m} Z, Z \leqq m\right) &= \tfrac{1}{2}(1 - \delta_{a_3 c_3}) G(m) \\
&\quad - \tfrac{1}{2}[T(m, a_3) + T(m, c_3)] - [S(m, a_3, b_3) + S(m, c_3, d_3)].
\end{aligned} \tag{3.14}
$$

The other two probabilities on the right side of (3.8) can be obtained from (3.14) by replacing m, a_3, b_3, c_3, d_3 by h, a_1, b_1, c_1, d_1 and k, a_2, b_2, c_2, d_2, respectively. Summing the expressions for these three probabilities gives (2.1). Equation (2.2) follows by noting that if h, k, and m are nonnegative or nonpositive, then

$$
\begin{aligned}
C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) &= \int_{-\infty}^{h} \int_{-\infty}^{k} \int_{-\infty}^{m} f(x, y, z; \rho_{12}, \rho_{13}, \rho_{23})\, dx\, dy\, dz \\
&= \int_{-\infty}^{h} \int_{-\infty}^{k} \int_{-m}^{\infty} f(x, y, z; \rho_{12}, -\rho_{13}, -\rho_{23})\, dx\, dy\, dz \\
&= \int_{-\infty}^{h} \int_{-\infty}^{k} \left(\int_{-\infty}^{\infty} - \int_{-\infty}^{-m}\right) f(x, y, z; \rho_{12}, -\rho_{13}, -\rho_{23})\, dx\, dy\, dz \\
&= B(h, k; \rho_{12}) - C(h, k, -m; \rho_{12}, -\rho_{13}, -\rho_{23}).
\end{aligned}
$$

The reader can verify that the familiar expression

$$
C(0, 0, 0; \rho_{12}, \rho_{13}, \rho_{23}) = \frac{1}{4\pi}(2\pi - \arccos \rho_{12} - \arccos \rho_{13} - \arccos \rho_{23})
$$

holds when $h = k = m = 0$ is substituted in (2.1).

If the G-function in the integrand of (3.7) is expanded in a Taylor series to three terms with remainder about the point $h(1 + a^2 + a^2(b/2)^2)^{1/2}$, then the following limited expansion can be shown to hold for $S(h, a, b)$.

$$
\begin{aligned}
S(h, a, b) &= \frac{1}{2\pi} G(h(1 + a^2 + a^2(b/2)^2)^{1/2}) \arctan (b/(1 + a^2 + a^2 b^2)^{1/2}) \\
&\quad + \frac{h}{2\pi} G'(h(1 + a^2 + a^2(b/2)^2)^{1/2}) \cdot \Delta_1(a, b) \\
&\quad + \frac{h^2}{2!\,2\pi} G''(h(1 + a^2 + a^2(b/2)^2)^{1/2}) \cdot \Delta_2(a, b) \\
&\quad + \frac{\theta h^3}{3!\,2\pi} \sup_{0 \leqq \xi \leqq 1} G'''(h(1 + a^2 + a^2 \xi^2)^{1/2}) \cdot \Delta_3(a, b),
\end{aligned} \tag{3.15}
$$

where $|\theta| \leqq 1$, and

$$\Delta_1(a, b) = \arctan b - \{1 + a^2 + a^2(b/2)^2\}^{1/2} \arctan(b/(1 + a^2 + a^2b^2)^{1/2}),$$

$$\begin{aligned}\Delta_2(a, b) = &\{2 + a^2 + a^2(b/2)^2\}\arctan(b/(1 + a^2 + a^2b^2)^{1/2}) \\ &- 2\{1 + a^2 + a^2(b/2)^2\}^{1/2} \arctan b \\ &+ a\{\log(ab + (1 + a^2 + a^2b^2)^{1/2}) - \tfrac{1}{2}\log(1 + a^2)\},\end{aligned}$$

$$\begin{aligned}\Delta_3(a, b) = &a^2b + \{4 + 3a^2 + 3a^2(b/2)^2\}\arctan b \\ &-\{1 + a^2 + a^2(b/2)^2\}^{3/2}\arctan(b/(1 + a^2 + a^2b^2)^{1/2}) \\ &- 3\{1 + a^2 + a^2(b/2)^2\}^{1/2}\{\arctan(b/(1 + a^2 + a^2b^2)^{1/2}) \\ &+ a[\log(ab + (1 + a^2 + a^2b^2)^{1/2}) - \tfrac{1}{2}\log(1 + a^2)]\}.\end{aligned}$$

If the first term of the series is used, the maximum error is one in the fourth decimal place for $h \leqq 2$ and one in the fifth decimal place for $h > 2$, and if the first three terms of this series are used, the maximum error encountered will be less than six in the sixth decimal place (note from (2.9) that the arc-tangent terms in the series can be read from the $h = \infty$ entries in the table).

4. Description of the table. The values of $S(h, a, b)$ given in the table were computed using a seven-point Gaussian quadrature formula on (3 7). The G-function in the integrand of (3.7) was approximated by a formula of C. Hastings (see [10], p. 187). A check of the computations was made for selected parameter values first by using an eight-point Gaussian quadrature formula with an improved method for evaluating $G(x)$ and second by using a sixteen-point Gaussian quadrature formula with the same improved method for evaluating $G(x)$. These two checks agreed with each other to nine decimal places and differed from the initially computed values by at most one in the eighth decimal place. These checks indicate that the tabulated values may occasionally be off by as much as 0.6 in the seventh decimal place because of rounding errors. Any number in the table whose last nonzero digit is a five is followed by a plus or minus sign to indicate that the number should be rounded up or down, respectively, when dropping the five.

The range of parameter values for which the S-function is tabulated was chosen so that outside the table $S(h, a, b)$ may be approximated by the first term of (3.15) with an error not exceeding five in the fifth decimal place.

The accuracy of linear interpolation in the table was checked empirically in the following way. Let Δh, Δa, Δb denote the intervals of tabulation on h, a, b, respectively. The check was performed by computing $S(h + \tfrac{1}{2}\Delta h, a + \tfrac{1}{2}\Delta a, b + \tfrac{1}{2}\Delta b)$ for a systematic selection of h, a, and b. Even though the errors found in this way are not necessarily the maximum errors in the various incremental cubes, it is felt that they are a reasonable approximation to these maximum errors. The errors found varied from about one to thirty in the fifth decimal place, which indicates that linear interpolation anywhere in the table should give an error of less than four or five in the fourth decimal place.

5. A numerical example. In [6] Plackett applies his reduction method to th computation of

$$D(-1.2, -1.0, 0.5; 0.7, 0.2, -0.4) = C(1.2, 1.0, -0.5; 0.7, 0.2, -0.4)$$
$$= B(1.2, 1.0; 0.7) - C(1.2, 1.0, 0.5; 0.7, -0.2, 0.4)$$

The numerical values of the constants defined by (2.3) are:

$$\begin{array}{lll} a_1 = 0.1867040 & b_1 = 4.0873367 & h = 1.2 \\ a_2 = 0.1091089 & b_2 = 10.5175180 & k = 1.0 \\ a_3 = 2.6536139 & b_3 = -0.4252646 & m = 0.5 \\ c_1 = 0.6293828 & c_2 = 0.7001401 & c_3 = 1.7457432 \\ d_1 = -0.7470863 & d_2 = 1.3079477 & d_3 = 1.3146897, \end{array}$$

and, therefore, by (2.1)

$$\begin{aligned} C(1.2, 1.0, 0.5; 0.7, -0.2, 0.4) &= \tfrac{1}{2}[G(1.2) + G(1) + G(0.5)] \\ &- \tfrac{1}{2}[T_1(1.2, 0.1867040) + T_2(1.2, 0.6293828) \\ &+ T_3(1, 0.1091089) + T_4(1, 0.7001401) \\ &+ T_5(0.5, 2.6536139) + T_6(0.5, 1.7457432)] \\ &- [S_1(1.2, 0.1867040, 4.0873367) + S_2(1.2, 0.6293828, -0.7470863) \\ &+ S_3(1, 0.1091089, 10.5175180) + S_4(1, 0.7001401, 1.3079477) \\ &+ S_5(0.5, 2.6536139, -0.4252646) + S_6(0.5, 1.7457432, 1.3146897)]. \end{aligned}$$

Tables of the G-function give $\frac{1}{2}[G(1.2) + G(1) + G(0.5)] = 1.2088688$, and the tables in [9] or [10] give

$$-\tfrac{1}{2}\sum T_i = -0.2025741, \qquad B(1.2, 1.0; 0.7) = 0.7940171.$$

Applying (2.5) to compute S_1, S_2, and S_3 and (2.4) to compute S_6, one finds

$$\begin{array}{lll} S_1 = 0.1808805, & S_2 = -0.0783075, & S_3 = 0.1927877, \\ S_4 = 0.1016940, & S_5 = -0.0204185, & S_6 = 0.0562510, \end{array}$$

and $\sum S_i = -0.4328872$, giving $C(1.2, 1.0, 0.5; 0.7, -0.2, 0.4) = 0.5734075$, and $D(-1.2, -1.0, 0.5; 0.7, 0.2, -0.4) = 0.2206096$.

If the bivariate probability $P(X > -1.0, Y > 0.5; \rho = -0.559714)$, incorrectly computed by Plackett as 0.587191, is given its correct value of 0.204267, then Plackett's answer is

$$D(-1.2, -1.0, 0.5; 0.7, 0.2, -0.4) = 0.220610,$$

and the answers agree to six decimal places.

6. Extension of method to higher dimensions. Equation (3.8) can be generalized to any number of dimensions giving

$$
\begin{aligned}
&P(X_1 \leqq u_1, X_2 \leqq u_2, \cdots, X_n \leqq u_n) \\
&\quad = P\left(X_1 \leqq u_1, X_2 \leqq \frac{u_2}{u_1} X_1, \cdots, X_n \leqq \frac{u_n}{u_1} X_1\right) \\
&\quad + P\left(X_1 \leqq \frac{u_1}{u_2} X_2, X_2 \leqq u_2, \cdots, X_n \leqq \frac{u_n}{u_2} X_2\right) \\
&\quad + \cdots \\
&\quad + P\left(X_1 \leqq \frac{u_1}{u_n} X_n, X_2 \leqq \frac{u_2}{u_n} X_n, \cdots, X_n \leqq u_n\right).
\end{aligned} \tag{6.1}
$$

provided all the u_i's are nonnegative (or nonpositive) and 0/0 is taken as one. Each term on the right side of (6.1) is expressible as an integral of a lower dimensional probability, for example,

$$
\begin{aligned}
&P\left(X_1 \leqq \frac{u_1}{u_n} X_n, X_2 \leqq \frac{u_2}{u_n} X_n, \cdots, X_n \leqq u_n\right) \\
&\quad = \int_{-\infty}^{u_n} P\left(X_1 \leqq \frac{u_1}{u_n} s, \cdots, X_{n-1} \leqq \frac{u_{n-1}}{u_n} s \,\middle|\, X_n = s\right) G'(s)\, ds.
\end{aligned} \tag{6.2}
$$

Since the three-dimensional normal distribution can be tabulated as a function of three variables, it follows by mathematical induction, using (6.1) and (6.2), that the n-dimensional normal distribution can be tabulated as a function of n variables.

As an example, consider the case $n = 4$. If $EX_i = 0$, $EX_i^2 = 1$, and $EX_iX_j = \rho_{ij}$ then the probability in the integrand of (6.2) can be expressed as

$$
\begin{aligned}
&P\left(X_1 \leqq \frac{u_1}{u_4} s, X_2 \leqq \frac{u_2}{u_4} s, X_3 \leqq \frac{u_3}{u_4} s \,\middle|\, X_4 = s\right) \\
&\quad = C(\alpha_{41} s, \alpha_{42} s, \alpha_{43} s; \dot\rho_{12}, \dot\rho_{13}, \dot\rho_{23}),
\end{aligned} \tag{6.3}
$$

where

$$
\alpha_{4i} = \frac{u_i - u_4 \rho_{i4}}{u_4(1 - \rho_{i4}^2)^{1/2}} \qquad \dot\rho_{ij} = \frac{\rho_{ij} - \rho_{i4}\rho_{j4}}{[(1 - \rho_{i4}^2)(1 - \rho_{j4}^2)]^{1/2}}.
$$

Therefore, (6.2) can be written as

$$
\begin{aligned}
&P\left(X_1 \leqq \frac{u_1}{u_n} X_n, X_2 \leqq \frac{u_2}{u_n} X_n, \cdots, X_n \leqq u_n\right) \\
&\quad = \int_{-\infty}^{u_n} C(\alpha_{41} s, \alpha_{42} s, \alpha_{43} s; \dot\rho_{12}, \dot\rho_{13}, \dot\rho_{23}) G'(s)\, ds.
\end{aligned}
$$

If the integrand of (6.4) is expressed by (2.10) and the result integrated, it is apparent that the left side of (6.1) can be expressed in terms of the G-, T-, and S-functions and integrals of the form

$$
R(h, a, b, c) = \int_{-\infty}^{h} S(as, b, c) G'(s)\, ds.
$$

TABLE

	m \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
a = 0	0.0	.0079314	.0157082	.0231934	.0302797	.0368959	.0430052	.0486000	.0536942	.0583156	.0625000
	0.1	.0085632	.0169595-	.0250409	.0326917	.0398349	.0464303	.0524713	.0579712	.0629607	.0674765-
	0.2	.0091887	.0181983	.0268700	.0350797	.0427446	.0498224	.0563041	.0622057	.0675597	.0724075-
	0.3	.0098018	.0194126	.0286629	.0374201	.0455968	.0531479	.0600610	.0663565	.0720677	.0772380
	0.4	.0103968	.0205910	.0304029	.0396920	.0483648	.0563731	.0637070	.0703847	.0764426	.0819277
	0.5	.0109685+	.0217233	.0320747	.0418746	.0510243	.0594730	.0672102	.0742550+	.0806161	.0864328
	0.6	.0115123	.0228004	.0336651	.0439503	.0535512	.0624218	.0705426	.0779368	.0846447	.0907184
	0.7	.0120245+	.0238148	.0351629	.0459063	.0559369	.0651990	.0736612	.0814013	.0884105	.0947545+
	0.8	.0125021	.0247607	.0365595-	.0477296	.0581586	.0677886	.0766077	.0846375+	.0919222	.0985181
	0.9	.0129431	.0256340	.0378488	.0494129	.0602097	.0701793	.0793034	.0876224	.0951640	.1019925-
	1.0	.0133460	.0264321	.0390273	.0509514	.0620843	.0723644	.0817788	.0903505	.0981270	.1051681
	1.1	.0137107	.0271543	.0400937	.0523436	.0637803	.0743417	.0840133	.0928194	.1008033	.1080417
	1.2	.0140374	.0278014	.0410491	.0535909	.0653006	.0761132	.0860153	.0950312	.1032104	.1105163
	1.3	.0143272	.0283754	.0418965+	.0546973	.0666437	.0776846	.0877911	.0969931	.1053412	.1127000
	1.4	.0145817	.0288794	.0426408	.0556689	.0678326	.0790645+	.0893505+	.0987160	.1072124	.1149054
	1.5	.0148030	.0293176	.0432878	.0565137	.0688620	.0802643	.0907061	.1002141	.1088394	.1166491
	∞	.0158628	.0314165-	.0463868	.0605595-	.0737918	.0860104	.0972001	.1073884	.1166311	.1250000
a = 0.1	0.0	.0078919	.0156293	.0230749	.0301218	.0366986	.0427685-	.0483239	.0533767	.0579603	.0621050
	0.1	.0085237	.0168805-	.0249224	.0325337	.0396375-	.0461940	.0521951	.0576557	.0626059	.0670844
	0.2	.0091491	.0181192	.0267513	.0349214	.0425469	.0495851	.0560274	.0618836	.0672012	.0720127
	0.3	.0097620	.0193331	.0285436	.0372614	.0453981	.0529084	.0597830	.0660383	.0717105	.0768422
	0.4	.0103567	.0205108	.0302825+	.0395315+	.0481642	.0561325+	.0634264	.0700641	.0760821	.0815273
	0.5	.0109278	.0216419	.0319526	.0417118	.0508208	.0592289	.0669255-	.0739298	.0802803	.0860266
	0.6	.0114708	.0227173	.0335405-	.0437847	.0533466	.0621727	.0702522	.0776049	.0842715-	.0903039
	0.7	.0119819	.0237296	.0350350+	.0457358	.0557238	.0649435-	.0733831	.0810637	.0880277	.0943232
	0.8	.0124582	.0246728	.0364276	.0475538	.0579388	.0675250+	.0763002	.0842862	.0915271	.0980793
	0.9	.0128975-	.0255428	.0377121	.0492307	.0599819	.0699061	.0789907	.0872584	.0947546	.1015378
	1.0	.0132987	.0263373	.0388851	.0507619	.0618475+	.0720803	.0814474	.0899720	.0977012	.1046952
	1.1	.0136613	.0270555+	.0399455+	.0521461	.0635339	.0740456	.0836679	.0924247	.1003614	.1075488
	1.2	.0139859	.0276983	.0408944	.0533848	.0650430	.0758041	.0856547	.0946193	.1027472	.1101018
	1.3	.0142734	.0282677	.0417351	.0544821	.0663798	.0773619	.0874147	.0965631	.1048576	.1123629
	1.4	.0145256	.0287671	.0424724	.0554444	.0675521	.0787280	.0889580	.0982676	.1067031	.1143452
	1.5	.0147446	.0292007	.0431125+	.0562799	.0685699	.0799139	.0902977	.0997470	.1083142	.1160658
	∞	.0157838	.0312585+	.0461499	.0602437	.0733971	.0855369	.0966478	.1067575-	.1159217	.1242121
a = 0.2	0.0	.0077769	.0153993	.0227302	.0296626	.0361251	.0420812	.0475234	.0524654	.0569355-	.0603693
	0.1	.0084086	.0166505-	.0245775+	.0320743	.0390638	.0455066	.0513943	.0567421	.0615802	.0659473
	0.2	.0090338	.0178887	.0264058	.0344611	.0419721	.0488964	.0552251	.0609743	.0661766	.0708734
	0.3	.0096462	.0191014	.0281964	.0367988	.0448204	.0522162	.0589766	.0651189	.0706776	.0756971
	0.4	.0102397	.0202769	.0299320	.0390645-	.0475810	.0554336	.0626123	.0691353	.0750393	.0803712
	0.5	.0108091	.0214046	.0315968	.0412378	.0502289	.0585196	.0660992	.0729871	.0792218	.0848531
	0.6	.0113496	.0224751	.0331773	.0433009	.0527424	.0614486	.0694086	.0766425-	.0831908	.0891057
	0.7	.0118575+	.0234808	.0346621	.0452390	.0551034	.0641999	.0725169	.0800754	.0869179	.0930987
	0.8	.0123297	.0244160	.0360426	.0470409	.0572984	.0667574	.0754059	.0832658	.0903813	.0968087
	0.9	.0127643	.0252766	.0373130	.0486988	.0593178	.0691100	.0780633	.0862001	.0935661	.1002199
	1.0	.0131601	.0260602	.0384697	.0502084	.0611563	.0712518	.0804821	.0888706	.0964642	.1033234
	1.1	.0135168	.0267665+	.0395122	.0515688	.0628129	.0731814	.0826610	.0912757	.0990740	.1061177
	1.2	.0138349	.0273965+	.0404420	.0527819	.0642901	.0749017	.0846032	.0931193	.1013995-	.1086071
	1.3	.0141158	.0279526	.0412627	.0538526	.0655936	.0764195+	.0863166	.0953100	.1034502	.1108019
	1.4	.0143612	.0284384	.0419795+	.0547877	.0667319	.0777448	.0878124	.0969602	.1052396	.1127166
	1.5	.0145733	.0288584	.0425992	.0555960	.0677157	.0788899	.0891045+	.0983855-	.1067848	.1143696
	∞	.0155537	.0307986	.0454604	.0593251	.0722502	.0841624	.0950467	.1019309	.1138709	.1219386

TABLE

a	m \ b	0.1	0.8
a = 0.3	0.0	.0075958	.0510441
	0.1	.0082275+	.0553204
	0.2	.0088524	.0595497
	0.3	.0094637	.0636865-
	0.4	.0100555-	.0676883
	0.5	.0106220	.0715169
	0.6	.0111584	.0751397
	0.7	.0116609	.0785299
	0.8	.0121266	.0816676
	0.9	.0125533	.0845398
	1.0	.0129402	.0871399
	1.1	.0132872	.0894679
	1.2	.0135950+	.0915293
	1.3	.0138651	.0933346
	1.4	.0140995+	.0948982
	1.5	.0143008	.0962377
	∞	.0151917	.1020882
a = 0.4	0.0	.0073624	.0492383
	0.1	.0079940	.0535140
	0.2	.0086184	.0577392
	0.3	.0092284	.0618651
	0.4	.0098175+	.0658465+
	0.5	.0103800	.0696431
	0.6	.0109108	.0732206
	0.7	.0114059	.0765520
	0.8	.0118624	.0796174
	0.9	.0122785+	.0824049
	1.0	.0126534	.0849097
	1.1	.0129872	.0871339
	1.2	.0132810	.0890856
	1.3	.0135367	.0907780
	1.4	.0137566	.0922282
	1.5	.0139436	.0934563
	∞	.0147248	.0984766
a = 0.5	0.0	.0070917	.0471792
	0.1	.0077232	.0514541
	0.2	.0083469	.0556740
	0.3	.0089552	.0597861
	0.4	.0095411	.0637415-
	0.5	.0100984	.0674972
	0.6	.0106220	.0710174
	0.7	.0111077	.0742746
	0.8	.0115528	.0772496
	0.9	.0119556	.0799320
	1.0	.0123155-	.0823194
	1.1	.0126331	.0844171
	1.2	.0129099	.0862365+
	1.3	.0131482	.0877943
	1.4	.0133508	.0891110
	1.5	.0135208	.0902097
	∞	.0141834	.0943583

	m \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
	0.0	.0067981	.0134463	.0198118	.0257918	.0313188	.0363597	.0409097	.0449854	.0486166	.0518406
	0.1	.0074295+	.0146968	.0216582	.0282023	.0342559	.0397830	.0447783	.0492594	.0532583	.0568153
	0.2	.0080524	.0159303	.0234794	.0305797	.0371525+	.0131588	.0485928	.0534729	.0578337	.0617181
	0.3	.0086586	.0171307	.0252515-	.0328926	.0399699	.0464415+	.0523010	.0575681	.0622792	.0664801
	0.4	.0092405-	.0182830	.0269522	.0351119	.0426725+	.0495894	.0558556	.0614918	.0665366	.0710386
	0.5	.0097916	.0193741	.0285623	.0372123	.0452293	.0525661	.0592152	.0651982	.0705558	.0753391
	0.6	.0103065+	.0203934	.0300659	.0391730	.0476149	.0553419	.0623460	.0686498	.0742959	.0793379
	0.7	.0107811	.0213326	.0314510	.0409782	.0498100	.0578943	.0652227	.0718186	.0777266	.0830024
a = 0.6	0.8	.0112126	.0221863	.0327094	.0426175+	.0518021	.0602039	.0678290	.0746868	.0808286	.0863123
	0.9	.0115996	.0229518	.0338374	.0440859	.0535851	.0622786	.0701572	.0772463	.0835934	.0892588
	1.0	.0119420	.0236289	.0348345+	.0453832	.0551589	.0641037	.0722080	.0794979	.0860227	.0918443
	1.1	.0122409	.0242198	.0357041	.0465136	.0565290	.0656908	.0739892	.0814509	.0881268	.0940803
	1.2	.0124983	.0247283	.0364522	.0474851	.0577053	.0670517	.0755145-	.0831210	.0899233	.0959866
	1.3	.0127170	.0251601	.0370869	.0483087	.0587013	.0682026	.0768025+	.0845291	.0914355-	.0975884
	1.4	.0129002	.0255218	.0376181	.0489974	.0595332	.0691624	.0778750-	.0856996	.0926903	.0989153
	1.5	.0130516	.0258207	.0380567	.0495653	.0602183	.0699517	.0787555-	.0866589	.0937169	.0999988
	∞	.0135962	.0268926	.0396236	.0515836	.0626377	.0727194	.0818195-	.0899707	.0972331	.1036813
	0.0	.0064941	.0128409	.0189104	.0246019	.0298500	.0346233	.0389184	.0427525+	.0461562	.0491670
	0.1	.0071254	.0140912	.0207563	.0270118	.0327364	.0380458	.0427860	.0470254	.0507967	.0541403
	0.2	.0077473	.0153228	.0225747	.0293855-	.0356783	.0414160	.0465938	.0512314	.0553636	.0590336
	0.3	.0083510	.0165182	.0243393	.0316884	.0384833	.0446838	.0502847	.0553067	.0597867	.0637708
	0.4	.0089282	.0176612	.0260261	.0338891	.0411627	.0478038	.0538069	.0591934	.0640026	.0682832
	0.5	.0094721	.0187379	.0276145+	.0359606	.0436834	.0507373	.0571161	.0628424	.0679574	.0725124
	0.6	.0099770	.0197370	.0290880	.0378811	.0460189	.0534531	.0601772	.0662146	.0716085-	.0764126
	0.7	.0104387	.0206504	.0304344	.0396350+	.0481500-	.0559290	.0629649	.0692822	.0749259	.0799520
a = 0.7	0.8	.0108546	.0214731	.0316465-	.0412126	.0500652	.0581517	.0654646	.0720292	.0778924	.0831125-
	0.9	.0112238	.0222031	.0327212	.0426104	.0517603	.0601165+	.0676713	.0744508	.0805035-	.0858897
	1.0	.0115467	.0228412	.0336601	.0438302	.0532378	.0618269	.0695893	.0765522	.0827654	.0882914
	1.1	.0118249	.0233907	.0344680	.0448787	.0545063	.0632931	.0712308	.0783473	.0846941	.0903353
	1.2	.0120611	.0238569	.0351528	.0457665+	.0555788	.0645307	.0726139	.0798571	.0863129	.0920473
	1.3	.0122586	.0242466	.0357246	.0465069	.0564719	.0655595-	.0737614	.0811070	.0876504	.0934587
	1.4	.0124213	.0245675-	.0361950-	.0471152	.0572043	.0664016	.0746987	.0821259	.0887381	.0946040
	1.5	.0125534	.0248277	.0365762	.0476073	.0577959	.0670804	.0754527	.0829435+	.0896090	.0955188
	∞	.0129882	.0256819	.0378208	.0492038	.0597001	.0692466	.0778367	.0855050+	.0923124	.0983340
	0.0	.0061894	.0122347	.0180090	.0234144	.0283879	.0328998	.0369480	.0405506	.0437383	.0465488
	0.1	.0068205-	.0134846	.0198545-	.0258237	.0313235+	.0363214	.0408145+	.0448222	.0483774	.0515206
	0.2	.0074413	.0147141	.0216696	.0281930	.0342100	.0396850-	.0446147	.0490195+	.0529346	.0564030
	0.3	.0080421	.0159038	.0234256	.0304845+	.0370007	.0429356	.0482856	.0530720	.0573320	.0611118
	0.4	.0086141	.0170361	.0250965-	.0326640	.0396535+	.0460239	.0517708	.0569165-	.0615005+	.0655716
	0.5	.0091497	.0180963	.0266602	.0347027	.0421332	.0489083	.0550229	.0605004	.0653823	.0697199
	0.6	.0096432	.0190727	.0280997	.0365779	.0444123	.0515566	.0580055-	.0637832	.0689334	.0735097
	0.7	.0100904	.0199573	.0294029	.0382744	.0464719	.0539471	.0606941	.0667382	.0721249	.0769102
a = 0.8	0.8	.0104892	.0207456	.0305635+	.0397837	.0483021	.0560683	.0630762	.0693520	.0749428	.0799072
	0.9	.0108389	.0214367	.0315801	.0411042	.0499012	.0579189	.0651507	.0716239	.0773874	.0825017
	1.0	.0111406	.0220326	.0324559	.0422404	.0512751	.0595059	.0669264	.0735646	.0794710	.0847082
	1.1	.0113967	.0225380	.0331980	.0432018	.0524357	.0608441	.0684204	.0751937	.0812161	.0865518
	1.2	.0116105+	.0229597	.0338164	.0440020	.0533998	.0619533	.0696561	.0765379	.0826522	.0880652
	1.3	.0117862	.0233059	.0343234	.0446568	.0541873	.0628574	.0706606	.0776278	.0838137	.0892858
	1.4	.0119281	.0235853	.0347322	.0451839	.0548197	.0635817	.0714634	.0784965-	.0847368	.0902534
	1.5	.0120409	.0238072	.0350564	.0456011	.0553193	.0641523	.0720941	.0791770	.0854580	.0910071
	∞	.0123787	.0244694	.0360180	.0468289	.0567759	.0657996	.0738960	.0811011	.0874767	.0930977

TABLE

	m \ b			
a = 0 9	0 0	0	(	C
	0 1	0	.(	C
	0 2	0	.(	C
	0 3	.0	(	C
	0.4	0	.(	C
	0.5	.0	.(	C
	0 6	.0	.(	C
	0 7	.0	.(	.(
	0.8	.0	.(	(
	0 9	0	(	.C
	1 0	0	.(	.C
	1.1	.0	.(	.C
	1.2	.0	C	.C
	1 3	0	.C	.0
	1 4	.0	.C	0
	1 5	0	.(	C
	∞	0	C	C
a = 1 0	0.0	.0	C	0
	0 1	0	C	0
	0.2	.0	.C	0
	0.3	0	0	0
	0 4	.0	0	0
	0.5	0	C	C
	0 6	.0	.C	C
	0.7	.0	C	0
	0 8	.0	C	.0
	0 9	.0	C	0
	1 0	.0	.0	0
	1.1	.0	.0	0
	1 2	.0	0	.0
	1 3	0	.0	0
	1 4	.0	C	.C
	1.5	.0	.C	.C
	∞	0.	.C	0
a = 1,1	0 0	0	.0	.0
	0 1	.0	.0	0
	0.2	.0	.0	.0
	0.3	.0	0	0
	0 4	.0	.C	C
	0 5	.0	0	0
	0,6	.0	.0	0
	0.7	.0	0	0
	0 8	.0	.0	.0
	0 9	.0	0	0
	1.0	0	0	0
	1 1	0	.0	0
	1 2	.0	0	0
	1,3	0	0	0
	1 4	0	0	0
	1 5	0	0	.0
	∞	0	0	0

TABLE

	m \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
a = 1.2	0.0	.0050726	.0100174	.0147226	.0191032	.0231063	.0267093	.0299142	.0327405+	.0352184	.0373829
	0.1	.0057028	.0112656	.0165655+	.0215091	.0260377	.0301253	.0337748	.0370054	.0398499	.0423462
	0.2	.0063179	.0124836	.0183634	.0238554	.0288953	.0334549	.0375348	.0411567	.0443553	.0471713
	0.3	.0069037	.0136432	.0200743	.0260868	.0316111	.0366157	.0411010	.0450828	.0486188	.0517314
	0.4	.0074481	.0147206	.0216627	.0281566	.0341270	.0395401	.0443955-	.0487171	.0525437	.0550217
	0.5	.0079419	.0156972	.0231013	.0300287	.0363993	.0421765-	.0473596	.0519736	.0560594	.0596659
	0.6	.0083790	.0165610	.0243723	.0316802	.0383939	.0444926	.0499572	.0548199	.0591235+	.0629198
	0.7	.0087566	.0173066	.0254678	.0331011	.0401172	.0464755+	.0521747	.0572421	.0617227	.0655707
	0.8	.0090749	.0179345-	.0263889	.0342932	.0415513	.04812[illegible]	.0540187	.0592492	.0638686	.0678335+
	0.9	.0093368	.0184504	.0271445+	.0352086	.0427268	.0494751	.0555126	.0608690	.0655936	.0697452
	1.0	.0095470	.0188642	.0277491	.0360471	.0436594	.0505411	.0566917	.0621422	.0669433	.0711575-
	1.1	.0097117	.0191879	.0282211	.0366529	.0443826	.0513646	.0575936	.0631171	.0679731	.07222[illegible]
	1.2	.0098377	.0194350-	.0285805+	.0371129	.0449295+	.0519645+	.0582783	.0638145+	.0687376	.0730225+
	1.3	.0099317	.0196191	.0288476	.0374534	.0453327	.0524395+	.0587748	.0643733	.0692910	.0735940
	1.4	.0100001	.0197529	.0290411	.0376992	.0456226	.0527651	.0591283	.0647481	.0696813	.0739956
	1.5	.0100487	.0198478	.0291780	.0378724	.0458258	.0529922	.0593737	.0650059	.0699498	.0742707
	∞	.0101452	.0200349	.0294451	.0382053	.0462126	.0534185+	.0598284	.0654811	.0701363	.0747658
a = 1.3	0.0	.0048308	.0095382	.0140142	.0181773	.0219768	.0253916	.0284245+	.0310950+	.0334325+	.0354713
	0.1	.0054608	.0107859	.0158564	.0205821	.0249068	.0288066	.0322834	.0353578	.0380617	.0404320
	0.2	.0060741	.0120003	.0176489	.0229212	.0277556	.0321249	.0360308	.0394949	.0425511	.0452393
	0.3	.0066553	.0131507	.0193460	.0251343	.0304493	.0352532	.0395650+	.0433914	.0467735+	.0497541
	0.4	.0071914	.0142114	.0209095+	.0271709	.0329230	.0381332	.0428021	.0469535-	.0506256	.0538638
	0.5	.0076728	.0151633	.0223111-	.0289939	.0351311	.0406967	.0456818	.0501142	.0540344	.0574904
	0.6	.0080937	.0159947	.0235337	.0305812	.0370551	.0429181	.0481699	.0528367	.0569611	.0605938
	0.7	.0084518	.0167015+	.0245714	.0319256	.0386777	.0447887	.0502531	.0551135+	.0593996	.0631695+
	0.8	.0087485-	.0172864	.0254284	.0330330	.0400104	.0463197	.0519608	.0569625-	.0613715+	.0652439
	0.9	.0089877	.0177573	.0261171	.0339205+	.0410746	.0475376	.0533035+	.0584206	.0629198	.0668653
	1.0	.0091755+	.0181265+	.0266557	.0346123	.0419011	.0484793	.0543477	.0595378	.0641004	.0680960
	1.1	.0093190	.0184082	.0270655-	.0351369	.0425252	.0491872	.0551242	.0603693	.0649749	.0690035-
	1.2	.0094258	.0186174	.0273689	.0355238	.0429835-	.0497045+	.0556888	.0609707	.0656043	.0696537
	1.3	.0095031	.0187685+	.0275875+	.0358015-	.0433108	.0500720	.0560878	.0613936	.0660448	.0701067
	1.4	.0095576	.0188749	.0277408	.0359953	.0435381	.0503259	.0563620	.0616826	.0663444	.0704137
	∞	.0096616	.0190765-	.0280285-	.0363545+	.0439535+	.0507832	.0568491	.0621900	.0668650+	.0709426
a = 1.4	0.0	.0046050-	.0090908	.0133534	.0173144	.0209255+	.0241670	.0270421	.0295701	.0317800	.0337049
	0.1	.0052347	.0103379	.0151947	.0197181	.0238542	.0275802	.0308989	.0338307	.0364066	.0386628
	0.2	.0058460	.0115485-	.0169814	.0220495+	.0266933	.0308871	.0346329	.0379524	.0408788	.0434509
	0.3	.0064223	.0126890	.0186637	.0242429	.0293614	.0339909	.0381329	.0418099	.0450575-	.0479173
	0.4	.0069495+	.0137320	.0202008	.0262444	.0317923	.0368136	.0413092	.0453029	.0488323	.0519418
	0.5	.0074179	.0146579	.0215636	.0280158	.0339393	.0393007	.0441004	.0483634	.0521295+	.0554459
	0.6	.0078219	.0154557	.0227359	.0295366	.0357778	.0414240	.0464755+	.0509585+	.0549150+	.0583949
	0.7	.0081602	.0161229	.0237146	.0308029	.0373041	.0431806	.0484329	.0530884	.0571915+	.0607947
	0.8	.0084352	.0166646	.0245074	.0318258	.0385326	.0445888	.0499952	.0547808	.0589920	.0626837
	0.9	.0086522	.0170914	.0251306	.0326272	.0394913	.0456829	.0512033	.0560830	.0603703	.0641230
	1.0	.0088185-	.0174179	.0256060	.0332363	.0402167	.0465068	.0521084	.0570534	.0613923	.0651848
	1.1	.0089422	.0176603	.0259578	.0336853	.0407490	.0471082	.0527654	.0577540	.0621263	.0659437
	1.2	.0090315-	.0178350-	.0262105+	.0340065-	.0411278	.0475338	.0532278	.0582443	.0626372	.0664695-
	1.3	.0090941	.0179572	.0263867	.0342292	.0413892	.0478258	.0535431	.0585769	.0629820	.0668227
	1.4	.0091367	.0180401	.0265058	.0343792	.0415641	.0480200	.0537517	.0587956	.0632077	.0670530
	∞	.0092099	.0181816	.0267068	.0346288	.0418510	.0483340	.0540842	.0591403	.0635600	.0674097

TABLE

a	m \ b	0.1
a = 1.5	0.0	.0043945-
	0.1	.0050239
	0.2	.0056332
	0.3	.0062042
	0.4	.0067221
	0.5	.0071769
	0.6	.0075636
	0.7	.0078817
	0.8	.0081352
	0.9	.0083306
	1.0	.0084765+
	1.1	.0085820
	1.2	.0086557
	1.3	.0087057
	∞	.0087890
a = 1.6	0.0	.0041986
	0.1	.0048277
	0.2	.0054348
	0.3	.0060002
	0.4	.0065084
	0.5	.0069491
	0.6	.0073180
	0.7	.0076160
	0.8	.0078484
	0.9	.0080231
	1.0	.0081500-
	1.1	.0082389
	1.2	.0082990
	1.3	.0083382
	∞	.0083972
a = 1.7	0.0	.0040164
	0.1	.0046452
	0.2	.0052499
	0.3	.0058095-
	0.4	.0063074
	0.5	.0067337
	0.6	.0070846
	0.7	.0073626
	0.8	.0075743
	0.9	.0077295-
	1.0	.0078389
	1.1	.0079130
	1.2	.0079613
	∞	.0080328

TABLE

	m \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
a = 1.8	0.0	.0038469	.0075907	.0111413	[illegible]	.0174220	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.1	.0014753	.0088351	.0125736	.0163102	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.2	.0050776	.0100276	.0147342	.0191255+	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.3	.0056310	.0111225-	.016352[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.4	.0061164	.0120853	.0177700	.0230710	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.5	.0065298	.0128930	.0189623	.0246174	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.6	.0068627	.0135541	.0195242	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.0477760	[illegible]
	0.7	.0071208	.0140620	.0206660	.0265137	.0324302	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.8	.0073128	.0144383	.0212144	.0275160	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.9	.0074495+	.0147066	.0216026	.0250105+	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	1.0	.0075430	.0148890	.0215655	.0283439	.0342436	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	1.1	.0076041	.0150030	.0220366	[illegible]	.0344547	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	1.2	.0076425-	.0150825-	.0221431	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	∞	.0076939	.0151813	.0222826	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
a = 1.9	0.0	.0036892	.0072783	.0106820	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.0252135+	[illegible]
	0.1	.0013172	.0085225-	.0125131	.0162312	.0196173	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.2	.0019169	.0097098	.0142700	.0185162	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.3	.0054639	.0107918	.0158643	.0205931	[illegible]	[illegible]	[illegible]	[illegible]	.0431150-	.0405021
	0.4	.0059403	.0117333	.0172500-	[illegible]	.0271015-	.0313502	[illegible]	[illegible]	.0444524	.0440160
	0.5	.0063366	.0125153	.0183979	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.6	.0066514	.0131354	.0193056	.0250511	.0303023	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	0.7	.0068902	.0136019	.0199905-	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.0477115+	[illegible]
	0.8	.0070632	.0139442	.0201835-	.0265603	.0321019	.0370757	.0414837	.0451637	.0487506	[illegible]
	0.9	.0071829	.0141782	.0208221	[illegible]	[illegible]	[illegible]	[illegible]	.0460203	[illegible]	.0524022
	1.0	.0072620	.0143324	.0210441	.0272713	.0329374	.0380115-	.0425000-	[illegible]	[illegible]	[illegible]
	1.1	.0073119	.0144294	.0211329	.0274451	.0331396	.0382350+	[illegible]	[illegible]	.0501207	.0531021
	∞	.0073784	.0145576	.0213640	.0276688	.0333946	.0385122	.0430307	[illegible]	.0501271	.0534118
a = 2.0	0.0	.0035423	.0069884	.0102545+	.0132786	.0160236	.0181755-	[illegible]	.0225316	.0241775-	.0256011
	0.1	.0011699	.0082312	.0120894	.0156738	.0189416	.0218759	.0244810	.0267751	.0287819	.0305107
	0.2	.0017669	.0094131	.0138332	.0179480	.0217032	.0250972	.0281153	.0307831	[illegible]	.0351674
	0.3	.0053072	.0101818	.0154080	.0199934	.0241991	.0279879	.0313676	[illegible]	.0369930	[illegible]
	0.4	.0057724	.0114008	.0167596	.0217536	.0263239	.0301455	[illegible]	[illegible]	.0402351	[illegible]
	0.5	.0061533	.0121523	.0178620	.0231803	.0280139	.0324258	.0363287	[illegible]	[illegible]	.0454538
	0.6	.0064501	.0127366	.0187165+	.0242816	.0293648	.0339377	.0380035-	.0415872	.0447273	[illegible]
	0.7	.0066701	.0131687	.0193460	.0250889	.0303273	.0350321	.0392071	[illegible]	[illegible]	[illegible]
	0.8	.0068251	.0134725+	.0197868	.0256509	.0309929	.0357833	.0400269	.0437526	.0470036	[illegible]
	0.9	.0069292	.0136757	.0200800	.0260225+	.0314297	.0362721	.0405563	.0443121	[illegible]	.0501262
	1.0	.0069955+	.0138049	.0202655+	.0262559	.0317019	.0365745-	.0408806	.0446521	[illegible]	.0507841
	1.1	.0070358	.0138830	.0203770	.0263951	.0319791	.0367516	.0410691	.0448483	[illegible]	.0509882
	∞	.0070846	.0139768	.0205091	.0265573	.0320471	[illegible]	.0412781	.0450632	.0483519	.0512082
a = 2.2	0.0	.0032775+	.0064051	.0094846	.0122783	.0148118	.0170727	.0190656	.0208071	.0223201	.0236303
	0.1	.0039042	.0077062	.0113169	.0146700	.0177255-	.0201679	.0229015-	.0250437	.0269197	.0285583
	0.2	.0044955-	.0088765+	.0130433	.0169211	.0204643	.0236546	.0264956	.0290059	.0312129	.0331481
	0.3	.0050217	.0099171	.0145700	.0189155-	.0228845+	.0264621	.0296514	.0324723	.0349513	.0371318
	0.4	.0054634	.0107895-	.0158580	.0205785+	.0248951	.0287842	.0322489	.0353104	.0380007	.0403572
	0.5	.0058134	.0114792	.0168685+	.0218840	.0264654	.0305876	.0342537	.0374965-	.0403207	.0427965+
	0.6	.0060748	.0119934	.0176190	.0228486	.0276188	.0319031	.0357053	.0390501	.0419749	.0445227
	0.7	.0062592	.0123548	.0181442	.0235197	.0284155+	.0328047	.0366920	.0401042	.0430809	.0456680
	0.8	.0063817	.0125944	.0184905-	.0239592	.0289332	.0333856	.0373221	.0407714	.0437753	.0463818
	0.9	.0064587	.0127442	.0187057	.0242303	.0292498	.0337375-	.0377003	.0411684	.0441854	.0468005-
	1.0	.0065042	.0128324	.0188317	.0243878	.0294319	.0339380	.0379139	.0413908	.0444134	.0470320
	∞	.0065550+	.0129302	.0189603	.0245567	.0296237	.0341454	.0381313	.0416142	.0446402	.0472607

TABLE

a	m \ b
a = 2.4	0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, ∞
a = 2.6	0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, ∞
a = 2.8	0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, ∞
a = 3.0	0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, ∞
a = 3.2	0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, ∞

TABLE

	m \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
a = 3.4	0.0	.0022346	.0044059	.0064592	.0083546	.0100696	.0115935-	.0129334	.0141006	.0151117	.0159847
	0.1	.0028544	.0056331	.0082708	.0107186	.0129477	.0149473	.0167210	.0182821	.0196494	.0208440
	0.2	.0034015-	.0067154	.0098653	.0127943	.0154680	.0178728	.0200117	.0218991	.0235568	.0250080
	0.3	.0038279	.0075569	.0111007	.0143947	.0173996	.0200995+	.0224975-	.0246095+	.0264594	.0280740
	0.4	.0041212	.0081339	.0119433	.0154784	.0186963	.0215797	.0241324	.0263723	.0283260	.0300239
	0.5	.0042994	.0084828	.0124491	.0161231	.0194592	.0224400	.0250705-	.0273703	.0293703	.0311019
	0.6	.0043949	.0086688	.0127165+	.0164600	.0198528	.0228778	.0255412	.0278652	.0298811	.0316237
	0.7	.0044401	.0087563	.0128410	.0166149	.0200311	.0230731	.0257483	.0280301	.0301007	.0318461
	∞	.0044691	.0088117	.0129184	.0167092	.0201372	.0231870	.0258669	.0282013	.0302233	.0319694
a = 3.6	0.0	.0021195+	.0041789	.0061261	.0079231	.0095478	.0109929	.0122623	.0133678	.0143251	.0151516
	0.1	.0027379	.0054033	.0079334	.0102815-	.0124199	.0143382	.0160400	.0175380	.0188502	.0199969
	0.2	.0032763	.0064682	.0095019	.0123227	.0148972	.0172124	.0192713	.0210879	.0226824	.0240781
	0.3	.0036846	.0072735+	.0106834	.0138519	.0167409	.0193351	.0216376	.0236641	.0254374	.0269839
	0.4	.0039541	.0078032	.0114559	.0148437	.0179252	.0206839	.0231238	.0252625+	.0271259	.0287434
	0.5	.0041090	.0081062	.0118943	.0154009	.0185824	.0214225+	.0239264	.0261138	.0280133	.0296569
	0.6	.0041866	.0082570	.0121103	.0156720	.0188977	.0217716	.0243000	.0265047	.0284157	.0300667
	∞	.0042390	.0083578	.0122522	.0158462	.0190957	.0219858	.0245246	.0267356	.0286503	.0303033
a = 3.8	0.0	.0020154	.0039734	.0058245+	.0075326	.0090766	.0104495+	.0116553	.0127052	.0136141	.0143987
	0.1	.0026322	.0051948	.0076273	.0098850+	.0119412	.0137859	.0154226	.0168635-	.0181259	.0192291
	0.2	.0031616	.0062416	.0091689	.0118904	.0143739	.0166070	.0185924	.0203437	.0218805+	.0232251
	0.3	.0035516	.0070105+	.0102961	.0133479	.0161292	.0186253	.0208391	.0227860	.0244882	.0259713
	0.4	.0037981	.0074945+	.0110010	.0142514	.0172057	.0198484	.0221834	.0242281	.0260078	.0275511
	0.5	.0039318	.0077557	.0113780	.0147291	.0177674	.0204774	.0228644	.0249480	.0267558	.0283190
	0.6	.0039940	.0078764	.0115505-	.0149447	.0180170	.0207524	.0231575+	.0252534	.0270692	.0286373
	∞	.0040307	.0079468	.0116491	.0150652	.0181532	.0208991	.0233107	.0254104	.0272283	.0287974
a = 4.0	0.0	.0019206	.0037865+	.0055504	.0071777	.0086484	.0099559	.0111040	.0121035-	.0129686	.0137152
	0.1	.0025359	.0050048	.0073485-	.0095238	.0115051	.0132829	.0148604	.0162493	.0174663	.0185302
	0.2	.0030560	.0060330	.0088622	.0114922	.0138920	.0160493	.0179669	.0196578	.0211412	.0224385+
	0.3	.0034276	.0067654	.0099350+	.0128781	.0155589	.0179633	.0200944	.0219670	.0236029	.0250269
	0.4	.0036520	.0072056	.0105752	.0136971	.0165326	.0190670	.0213043	.0232616	.0249636	.0264381
	0.5	.0037665-	.0074289	.0108969	.0141035+	.0170088	.0195983	.0218775+	.0238655-	.0255892	.0270786
	∞	.0038413	.0075731	.0111008	.0143554	.0172968	.0199118	.0222081	.0242069	.0259372	.0274304
a = 4.2	0.0	.0018342	.0036160	.0053002	.0068539	.0082578	.0095057	.0106013	.0115548	.0123801	.0130922
	0.1	.0024478	.0048310	.0070934	.0091934	.0111063	.0128228	.0143462	.0156876	.0168632	.0178910
	0.2	.0029583	.0058400	.0085784	.0111238	.0134459	.0155330	.0173877	.0190227	.0204564	.0217097
	0.3	.0033116	.0065359	.0095971	.0124383	.0150250+	.0173436	.0193972	.0212003	.0227741	.0241428
	0.4	.0035149	.0069345-	.0101758	.0131772	.0159015+	.0183346	.0204807	.0223567	.0239865-	.0253973
	0.5	.0036122	.0071240	.0104482	.0135203	.0163022	.0187800	.0209595-	.0228593	.0245056	.0259275+
	∞	.0036684	.0072320	.0106005-	.0137077	.0165155+	.0190113	.0212025+	.0231096	.0247602	.0261844
a = 4.4	0.0	.0017550	.0034598	.0050711	.0065573	.0079001	.0090934	.0101410	.0110527	.0118415+	.0125222
	0.1	.0023669	.0046714	.0068591	.0088900	.0107400	.0124004	.0138740	.0151719	.0163095-	.0173041
	0.2	.0028675+	.0056607	.0083148	.0107814	.0130313	.0150530	.0168491	.0184318	.0198192	.0210314
	0.3	.0032026	.0063204	.0092797	.0120252	.0145235+	.0167616	.0187423	.0204801	.0219956	.0233125+
	0.4	.0033860	.0066796	.0098004	.0126888	.0153088	.0176471	.0197080	.0215080	.0230707	.0244224
	0.5	.0034681	.0068393	.0100294	.0129762	.0156432	.0180175-	.0201047	.0219231	.0234982	.0248581
	∞	.0035100	.0069196	.0101422	.0131146	.0158001	.0181869	.0202821	.0221053	.0236831	.0250444
a = 4.6	0.0	.0016822	.0033162	.0048605+	.0062847	.0075714	.0087148	.0097183	.0105915-	.0113470	.0119988
	0.1	.0022923	.0045242	.0066432	.0086103	.0104025-	.0120110	.0134389	.0146966	.0157991	.0167631
	0.2	.0027829	.0054934	.0080688	.0104619	.0126444	.0146049	.0163462	.0178800	.0192239	.0203976
	0.3	.0030999	.0061174	.0089806	.0116360	.0140511	.0162132	.0181254	.0198018	.0212625+	.0225308
	0.4	.0032646	.0064396	.0094471	.0122292	.0147513	.0170007	.0189819	.0207111	.0222112	.0235081
	0.5	.0033334	.0065732	.0096380	.0124681	.0150283	.0173064	.0193081	.0210514	.0225607	.0238634
	∞	.0033644	.0066325-	.0097211	.0125695-	.0151428	.0174295+	.0194366	.0211829	.0226940	.0239976

TABLE

	m \ b
a = 4.8	0.0
	0.1
	0.2
	0.3
	0.4
	∞
a = 5.0	0.0
	0.1
	0.2
	0.3
	0.4
	∞
a = 5.5	0.0
	0.1
	0.2
	0.3
	0.4
	∞
a = 6.0	0.0
	0.1
	0.2
	0.3
	∞
a = 6.5	0.0
	0.1
	0.2
	0.3
	∞
a = 7.0	0.0
	0.1
	0.2
	0.3
	∞
a = 7.5	0.0
	0.1
	0.2
	0.3
	∞
a = 8.0	0.0
	0.1
	0.2
	0.3
	∞

REFERENCES

H. RUBEN, "On the moments of order statistics in samples from normal populations," *Biometrika*, Vol. 41 (1954), pp. 200–227.

F. N. DAVID, "A note on the evaluation of the multivariate normal integral," *Biometrika*, Vol. 40 (1953), pp. 458–459.

D. TEICHROEW, "Probabilities associated with order statistics in samples from two normal populations with equal variances," Chemical Corps Engineering Agency, Army Chemical Center, Maryland, December 1955.

D. B. OWEN, unpublished tables submitted to members of the IMS Committee on Mathematical Tables.

M. G. KENDALL, "Proof of relations connected with the tetrachoric series and its generalization," *Biometrika*, Vol. 32 (1941), pp. 196–198.

R. L. PLACKETT, "A reduction formula for normal multivariate integrals," *Biometrika*, Vol. 41 (1954), pp. 351–360.

S. C. DAS, "The numerical evaluation of a class of integrals. II," *Proc. Cambridge Philos. Soc.*, Vol. 52 (1956), pp. 230–233.

D. B. OWEN, "The bivariate normal probability distribution," Office of Technical Services, Department of Commerce, Washington, D. C., 1957.

D. B. OWEN, "Tables for computing bivariate normal probabilities," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1075–1090.

C. HASTINGS, *Approximations for Digital Computers*, Princeton Univ. Press, 1955.

P. N. SOMERVILLE, "Some problems of optimum sampling," *Biometrika*, Vol. 41 (1954), pp. 420–429.

C. NICHOLSON, "The probability integral for two dimensions," *Biometrika*, Vol. 33 (1943), pp. 59–72.

NATIONAL BUREAU OF STANDARDS, *Tables of the Bivariate Normal Distribution Function and Related Functions*, (to be Published).

KARL PEARSON, *Tables for Statisticians and Biometricians*, Part II, Cambridge University Press, 1931.

ADMISSIBLE AND MINIMAX INTEGER-VALUED ESTIMATORS OF AN INTEGER-VALUED PARAMETER[1]

By D. S. Robson

Cornell University

1. Summary. The decision problem considered here is that of deciding which element of a finite parametric family of probability distributions $p(x, \mu)$ represents the true distribution of the statistic X. It is assumed that $p(x, \mu)$ satisfies certain regularity conditions which essentially require that the parameter μ be integer-valued with known bounds and that $p(x, \mu_1)/p(x, \mu_0)$ be an increasing function of x whenever $\mu_0 < \mu_1$. Complete classes are characterized for various loss functions $W(\mu, \alpha)$ which are convex functions of the decision α for each fixed value of μ. Minimax procedures are considered for the case $W(\mu, \alpha) = | \alpha - \mu |^k$.

2. Introduction. The problem of estimating an integer-valued parameter is viewed as a special case of Wald's general statistical decision problem. The chance variable X *is known to be distributed over the sample space* M *according* to a probability distribution $p(x, \mu)$ depending upon a single unknown integer-valued parameter μ with the known bounds $0 \leqq \mu \leqq N$. The statistician is required to make one of $N + 1$ decisions, corresponding to the $N + 1$ different possible values of μ, on the basis of a single observed value of X. A decision function δ therefore has the form

$$\delta(x) = (\delta_0(x), \delta_1(x), \cdots, \delta_N(x)) \tag{1}$$

where $\delta_\alpha(x) \geqq 0$ for $\alpha = 0, 1, \cdots, N$ and $\sum_{\alpha=0}^{N} \delta_\alpha(x) = 1$ for all x in M, with the interpretation that when the procedure δ is used and the observed value of X is x_0 then the decision that the true distribution of X is $p(x, \alpha)$ is to be made with probability $\delta_\alpha(x_0)$, $\alpha = 0, 1, \cdots, N$. The loss associated with the decision α when the true value of the parameter is μ is expressed by a loss function $W(\mu, \alpha)$ which, for each fixed value of μ, is a convex function of α with $W(\mu, \mu) = 0$ and, for α between μ and β, $W(\mu, \alpha) < W(\mu, \beta)$.

The following regularity conditions are imposed upon the function $p(x, \mu)$.

Condition 1. $p(y, \mu)p(x, v) < p(x, \mu)p(y, v)$ *if and only if* $p(y, \mu)p(x, v)$ *and* $p(x, \mu)p(y, v)$ *are not both zero and* $x < y$, $\mu < v$.

Condition 2. *If* $p(x, v) = 0$ *for all* x *in* M *then* $p(x, \mu) = 0$ *for all* x *in* M *either for every* $\mu \leqq v$ *or for every* $\mu \geqq v$.

Condition 3. *If* $M = (x_0, x_1, \cdots, x_n)$, $x_{i-1} < x_i$, *then for every* i, $0 < i \leqq n$, *there exists an integer* μ_i *such that* $p(x_{i-1}, \mu_i) > 0$ *and* $p(x_i, \mu_i) > 0$.

Conditions 1 and 2 are essentially a more precise way of saying that the likelihood ratio $p(x, v) / p(x, \mu)$ is a strictly increasing function of x whenever $\mu < v$.

Received October 5, 1955; revised January 10, 1958.

[1] Part of a doctoral thesis presented to the Faculty of the Graduate School of Cornell University.

A simple but useful consequence of Condition 1 is the following

LEMMA 1. *If the distribution* $p(x, \mu)$ *satisfies Condition* 1 *and if* $p(y, \alpha) = 0$ *and if there exists a pair* z, β *such that* $p(z, \alpha) > 0$ *and* $p(y, \beta) > 0$ *then either*

(i) $p(y, \mu) = 0$ *for all* $\mu \leqq \alpha$ *and* $p(x, \alpha) = 0$ *for all* $x \geqq y$, *or*

(ii) $p(y, \mu) = 0$ *for all* $\mu \geqq \alpha$ *and* $p(x, \alpha) = 0$ *for all* $x \leqq y$.

3. A Karlin-Rubin Complete Class Theorem. A general approach to decision problems involving distributions with a monotone likelihood ratio has been developed by H. Rubin [1] and S. Karlin and H. Rubin [2]. Since the finite action problem posed here represents a special case of the Karlin-Rubin problem, a direct application of their results concerning completeness of the class of monotone decision procedures gives the following

THEOREM 1. *Let C be the class of decision functions such that*

(i) *for every* x *in* M *there exists an integer* α_x *such that* $\delta_{\alpha_x}(x) + \delta_{\alpha_x+1}(x) = 1$

(ii) $\delta_\alpha(x) > 0$ *only if* $p(x, \alpha) > 0$

(iii) if $x < y$ then $\bar{\alpha}_x \equiv \alpha_x\delta_{\alpha_x}(x) + (\alpha_x + 1)\delta_{\alpha_x+1}(x) \leqq \alpha_y$

If $p(x, \mu)$ *satisfies Conditions* 1 *and* 2 *and if, for each fixed* μ, *the loss function* $W(\mu, \alpha)$ *is a convex function of* α *with* $W(\mu, \mu) = 0$ *and, for* α *between* μ *and* β, $W(\mu, \alpha) < W(\mu, \beta)$ *then the class C is complete.*

The theorem remains valid under weaker conditions on the loss function[2]; however, in what follows only convex loss functions are considered.

4. Admissible procedures when $W(\mu, \alpha) = |\alpha - \mu|^k$ **for large** k. The class C may, under the hypotheses of Theorem 1, contain inadmissible procedures. This is effectively demonstrated by the special case where $W(\mu, \alpha)$ is a convex function of $|\alpha - \mu|$ and increases very rapidly with $|\alpha - \mu|$. $W(\mu, \alpha) = |\alpha - \mu|^k$ is one example of such a loss function and, clearly, any convex function $W(|\alpha - \mu|)$ with $W(0) = 0$ can be dominated by $K|\alpha - \mu|^k$ by choosing the constants K and k sufficiently large. The most stringent requirements for admissibility are then encountered when the range of x for which $p(x, \mu) > 0$ is independent of μ; in particular,

THEOREM 2. *If* $p(x, \mu)$ *satisfies Condition* 1 *and* $p(x, \mu) > 0$ *for all integer pairs* (x, μ) *such that* $0 \leqq x \leqq n$, $0 \leqq \mu \leqq N$, *and* $p(x, \mu) = 0$ *otherwise, then there exists an integer* $k_p > 0$ *such that if* $W(\mu, \alpha) = |\alpha - \mu|^k$ *and* $k \geqq k_p$ *then every admissible procedure is of the form*

$$\delta_\alpha(x) = 1 \text{ for } x < y$$

$$\delta_\alpha(y) + \delta_{\alpha+1}(y) = 1$$

$$\delta_{\alpha+1}(x) = 1 \text{ for } x > y$$

where $0 \leqq y \leqq n, 0 \leqq \alpha \leqq N$.

[2] The author proved the theorem as it is stated and a referee pointed out that the Karlin-Rubin theorem for the finite action problem includes this result.

PROOF OF THEOREM 2. The conclusion is obtained by showing that, under the hypotheses specified, every Bayes solution has this form when k is sufficiently large. Let $\xi = (\xi_0, \xi_1, \cdots, \xi_N)$ be an a priori distribution on the parameter space and let $r(\xi, \delta)$ be the integrated risk of the procedure δ. Then δ^ξ is said to be a Bayes solution relative to ξ if $\inf_\delta r(\xi, \delta) = r(\xi, \delta^\xi)$. For every ξ, however, there exists a non-randomized Bayes solution; consequently, if

$$r_x(\xi, \alpha) = \sum_\mu |\alpha - \mu|^k p(x, \mu)\xi_\mu$$

then

$$r(\xi, \delta^\xi) = \sum_x \inf_\alpha r_x(\xi, \alpha).$$

The function $r_x(\xi, \alpha)$ is seen to have the following properties:

I. If $r_x(\xi, \alpha) \leqq r_x(\xi, \alpha + 1)$ then $r_x(\xi, \alpha + \beta) < r_x(\xi, \alpha + \beta + 1)$ for all $\beta > 0$

II. If $r_y(\xi, \alpha) \leqq r_y(\xi, \alpha + 1)$ then $r_x(\xi, \alpha) < r_x(\xi, \alpha + 1)$ for all $x \leqq y$

III: If $r_0(\xi, \alpha - 1) \leqq r_0(\xi, \alpha)$ then $r_n(\xi, \alpha) < r_n(\xi, \alpha + 1)$ for all $k \geqq k_p$.

Let $\Delta_k(\alpha) = (\alpha + 1)^k - \alpha^k$ then $r_x(\xi, \alpha) \leqq r_x(\xi, \alpha + 1)$ is equivalent to

$$\sum_{\mu=\alpha+1}^{N} \Delta_k(\mu - \alpha - 1)p(x, \mu)\xi_\mu \leqq \sum_{\mu=0}^{\alpha} \Delta_k(\alpha - \mu)p(x, \mu)\xi_\mu.$$

Then property I follows from

$$\begin{aligned}\sum_{\mu=\alpha+\beta+1}^{N} \Delta_k(\mu - \alpha - \beta - 1)p(x, \mu)\xi_\mu &\leqq \sum_{\mu=\alpha+1}^{N} \Delta_k(\mu - \alpha - 1)p(x, \mu)\xi_\mu \\ &\leqq \sum_{\mu=0}^{\alpha} \Delta_k(\alpha - \mu)p(x, \mu)\xi_\mu \leqq \sum_{\mu=0}^{\alpha+\beta} \Delta_k(\alpha + \beta - \mu)p(x, \mu)\xi_\mu\end{aligned} \tag{2}$$

for all β, $0 < \beta \leqq N - \alpha - 1$. Since $p(x, \mu) > 0$, $0 \leqq x \leqq n$, $0 \leqq \mu \leqq N$, then either the first or last inequality (or both) of (2) is strict for $k > 1$. Property II is a direct result of the restrictions upon $p(x, \mu)$, for if $r_y(\xi, \alpha) \leqq r_y(\xi, \alpha + 1)$ then, since $p(x, \mu)$ satisfies Condition 1 and $p(x, \mu) > 0$, $0 \leqq x \leqq n$, $0 \leqq \mu \leqq N$,

$$\begin{aligned}\sum_{\mu=\alpha+1}^{N} \Delta_k(\mu - \alpha - 1) \frac{p(x, \mu)}{p(x, \alpha)} \xi_\mu &\leqq \sum_{\mu=\alpha+1}^{N} \Delta_k(\mu - \alpha - 1) \frac{p(y, \mu)}{p(y, \alpha)} \xi_\mu \\ &\leqq \sum_{\mu=0}^{\alpha} \Delta_k(\alpha - \mu)\frac{p(y, \mu)}{p(y, \alpha)} \xi_\mu \leqq \sum_{\mu=0}^{\alpha} \Delta_k(\alpha - \mu) \frac{p(x, \mu)}{p(x, \alpha)} \xi_\mu\end{aligned} \tag{3}$$

for all $x \leqq y$, with either the first or last inequality (or both) of (3) being strict for all k. Property III is derived by noting that if $r_0(\xi, \alpha - 1) \leqq r_0(\xi, \alpha)$ then

$$\Delta_k(\alpha - 1)\xi_0 \geqq - \sum_{\mu=1}^{\alpha-1} \Delta_k(\alpha - 1 - \mu) \frac{p(0, \mu)}{p(0, 0)} \xi_\mu + \sum_{\mu=\alpha}^{N} \Delta_k(\mu - \alpha) \frac{p(0, \mu)}{p(0, 0)} \xi_\mu$$

and

$$
(4)\quad
\begin{aligned}
&\frac{\Delta_k(\alpha-1)}{p(n,0)}(r_n(\xi,\alpha+1)-r_n(\xi,\alpha)) \\
&\geqq \Delta_k(\alpha)\left(-\sum_{\mu=1}^{\alpha-1}\Delta_k(\alpha-1-\mu)\frac{p(0,\mu)}{p(0,0)}\xi_\mu+\sum_{\mu=\alpha}^{N}\Delta_k(\mu-\alpha)\frac{p(0,\mu)}{p(0,0)}\xi_\mu\right) \\
&+\Delta_k(\alpha-1)\left(\sum_{\mu=1}^{\alpha}\Delta_k(\alpha-\mu)\frac{p(n,\mu)}{p(n,0)}\xi_\mu-\sum_{\mu=\alpha+1}^{N}\Delta_k(\mu-\alpha-1)\frac{p(n,\mu)}{p(n,0)}\xi_\mu\right) \\
&=\sum_{\mu=1}^{\alpha-1}\left(\Delta_k(\alpha-1)\Delta_k(\alpha-\mu)\frac{p(n,\mu)}{p(n,0)}-\Delta_k(\alpha)\Delta_k(\alpha-\mu-1)\frac{p(0,\mu)}{p(0,0)}\right)\xi_\mu \\
&\qquad+\left(\Delta_k(\alpha-1)\frac{p(n,\alpha)}{p(n,0)}+\Delta_k(\alpha)\frac{p(0,\alpha)}{p(0,0)}\right)\xi_\alpha \\
&+\sum_{\mu=\alpha+1}^{N}\left(\Delta_k(\alpha)\Delta_k(\mu-\alpha)\frac{p(0,\mu)}{p(0,0)}-\Delta_k(\alpha-1)\Delta_k(\mu-\alpha-1)\frac{p(n,\mu)}{p(n,0)}\right)\xi_\mu .
\end{aligned}
$$

For $0 \leqq \mu \leqq \alpha - 1$, $\Delta_k(\alpha-1)\,\Delta_k(\alpha-\mu) \geqq \Delta_k(\alpha)\Delta_k(\alpha-\mu-1)$ and, by Condition 1, $(p(n,\mu)/p(n,0)) > p(0,\mu)/p(0,0))$ so the coefficient of ξ_μ in (4) is positive for all $\mu \leqq \alpha$. And since, for $\mu > \alpha$, $(\Delta_k(\alpha)\Delta_k(\mu-\alpha)/\Delta_k(\alpha-1)$ $\Delta_k(\mu-\alpha-1))$ can be made arbitrarily large by choosing k sufficiently large then $k_p(\mu, \alpha)$ may be defined as the smallest integer k such that

$$\frac{\Delta_k(\alpha)\Delta_k(\mu-\alpha)}{\Delta_k(\alpha-1)\Delta_k(\mu-\alpha-1)} > \frac{p(n,\mu)p(0,0)}{p(n,0)p(0,\mu)}.$$

Hence, for $k \geqq \max_\mu k_p(\mu, \alpha)$ the coefficient of ξ_μ in (4) is positive for all $\mu > \alpha$, and property III is thus established by taking

$$k_p = \max_{\substack{0 \leqq \alpha \leqq N \\ \alpha \leqq \mu \leqq N}} k_p(\mu, \alpha)$$

Now let α_x^ξ be an integer such that $\min_{0 \leqq \alpha \leqq N} r_x(\xi, \alpha) = r_x(\xi, \alpha_x^\xi)$. Then for $x < y$, $\alpha_x^\xi \leqq \alpha_y^\xi$. For suppose $x < y$ and $\alpha_y^\xi < \alpha_x^\xi$; since $r_y(\xi, \alpha_y^\xi) \leqq r_y(\xi, \alpha_y^\xi + 1)$ then, by II, $r_x(\xi, \alpha_y^\xi) < r_x(\xi, \alpha_y^\xi + 1)$ and then I implies the contradiction $r_x(\xi, \alpha_x^\xi - 1) < r_x(\xi, \alpha_x^\xi)$. For $k \geqq k_p$, III gives $r_n(\xi, \alpha_0^\xi + 1) < r_n(\xi, \alpha_0^\xi + 2)$ which implies by II, that $r_x(\xi, \alpha_0^\xi + 1) < r_x(\xi, \alpha_0^\xi + 2)$ for all $x \leqq n$ and this implies, by I, that $r_x(\xi, \alpha_0^\xi + \beta) < r_x(\xi, \alpha_0^\xi + \beta + 1)$ for all $\beta > 1$. Hence, $\alpha_0^\xi \leqq \alpha_x^\xi \leqq \alpha_0^\xi + 1$ for all x. If there exists a value of x such that $\alpha_x^\xi = \alpha_0^\xi + 1$ let y be the least such x, then $0 \leqq y \leqq n$ and

$$\alpha_x^\xi = \begin{cases} \alpha_0^\xi \text{ for } x < y \\ \alpha_0^\xi + 1 \text{ for } x \geqq y; \end{cases}$$

in this case, randomized Bayes solutions exist and are of the form

$$\delta_{\alpha_0^\xi}(x) = 1 \text{ for } x < y$$

$$\delta_{\alpha_0^\xi}(y) + \delta_{\alpha_0^\xi+1}(y) = 1$$

$$\delta_{\alpha_0^\xi+1}(x) = 1 \text{ for } x > y.$$

Since every admissible procedure is a Bayes solution this completes the proof of Theorem 2.

The distribution

$$(5)\qquad p(x,\mu) = \binom{n}{x}\left(\frac{\mu}{N}\right)^x\left(1-\frac{\mu}{N}\right)^{n-x}, \qquad 0 \leqq x \leqq n, 0 \leqq \mu \leqq N$$

satisfies the hypotheses of Theorem 2 for $0 < x < n$, and since Theorem 1 applies for $x = 0, n$ then

COROLLARY. *If $p(x, \mu)$ is the distribution* (5) *then when $k \geqq k_p$ a procedure δ is admissible only if*

$$\begin{aligned} \delta_\alpha(0) + \delta_{\alpha+1}(0) &= 1 \\ \delta_\beta(x) &= 1 \textit{ for } 0 < x < y \\ \delta_\beta(y) + \delta_{\beta+1}(y) &= 1 \\ \delta_{\beta+1}(x) &= 1 \textit{ for } y < x < n \\ \delta_\gamma(n) + \delta_{\gamma+1}(n) &= 1 \end{aligned}$$

where the integers α, β, γ satisfy $0 \leqq \alpha < \beta < \gamma \leqq N$.

5. Admissible procedures when $W(\mu, \alpha) = |\alpha - \mu|$. If C_k denotes the class of procedures which are admissible when $W(\mu, \alpha) = |\alpha - \mu|^k$ then C_k is contained in the class C of Theorem 1. As demonstrated by Theorem 2, however, when k is sufficiently large the class C_k may reduce to a collection of procedures which virtually designate the same decision for all values of x, so in this case little significance could be attached to the mere fact that a procedure belonged to the class C. Since $W(\mu, \alpha) = |\alpha - \mu|^k$ is a conventional type of loss function for estimation problems Theorem 2 therefore raises a question of the practical importance of the class C; hence, it is of special interest that

THEOREM 3. *If $p(x, \mu)$ satisfies Conditions* 1 *and* 2 *and if the sample space M is finite then the class C_1 of procedures which are admissible relative to $W(\mu, \alpha) = |\alpha - \mu|$ is the class C itself.*

PROOF OF THEOREM 3. If a member δ of C is inadmissible then, since C is a complete class, there exists a member δ' of C which is better than δ. Then for all possible μ

$$(6)\qquad r(\mu, \delta) - r(\mu, \delta') = \sum_x p(x, \mu)\,(|\bar{\alpha}_x - \mu| - |\bar{\alpha}'_x - \mu|) \geqq 0.$$

Theorem 3 is proved by showing that (6) cannot hold for all possible μ; and, in particular, that there exists x in M such that either

$$r(\alpha_x, \delta) > r(\alpha_x, \delta') \quad \text{or} \quad r(\alpha_x + 1, \delta) > r(\alpha_x + 1, \delta').$$

Only the ordering of the sample space is pertinent so, without loss of generality, let $M = (0, 1, \cdots, n)$ and let $|\bar{\alpha}_x - \bar{\alpha}'_x| = A_x > 0$ for all x in M. Let

$$d = \min_x(x \mid \bar\alpha_x < \bar\alpha_x')$$
$$e = \max_x(x \mid \bar\alpha_x > \bar\alpha_x')$$
$$Y = (x \mid d < x < e \text{ and } \bar\alpha_{x-1}' < \bar\alpha_{x-1} \leqq \bar\alpha_x < \bar\alpha_x')$$
$$= (y_1, y_2, \cdots, y_{m-1}), y_i < y_j \text{ for } i < j$$
$$Z = (x \mid d < x < e \text{ and } \bar\alpha_{x-1} < \bar\alpha_{x-1}' \leqq \bar\alpha_x' < \bar\alpha_x)$$
$$= (z_1, z_2, \cdots, z_m), z_i < z_j \text{ for } i < j$$

then let $y_0 = d$, $y_m = e$, and

$$u_{2i} = v_{2i-1} + 1 = y_i \text{ for } i = 0, 1, \cdots, m$$
$$u_{2i+1} = v_{2i} + 1 = z_{i+1} \text{ for } i = 0, 1, \cdots, m-1$$
$$\mu_{2i} = \alpha_{u_{2i}} \text{ for } i = 0, 1, \cdots, m-1$$
$$\mu_{2i+1} = \alpha_{v_{2i+1}} + 1 \text{ (or } \alpha_{v_{2i+1}} \text{ if } \alpha_{v_{2i+1}} = \bar\alpha_{v_{2i+1}}) \qquad \text{for } i = 0, 1, \cdots, m-1$$

Since δ and δ' are in C then $\bar\alpha_x \leqq \bar\alpha_y$ and $\bar\alpha_x' \leqq \bar\alpha_y'$ for all $x < y$ so

$$u_{2i} \leqq v_{2i} < u_{2i+1} \leqq v_{2i+1} \qquad \text{for } i = 0, 1, \cdots, m-1$$
$$\mu_0 < \mu_{2i-1} \leqq \mu_{2i} < \mu_{2i+1} \qquad \text{for } i = 1, 2, \cdots, m-1$$

and for $2k-1 \leqq q \leqq 2k$

$$r(\mu_q, \delta) - r(\mu_q, \delta') = -\sum_{x=0}^{u_0-1} p(x, \mu_q)A_x + \sum_{i=0}^{2k-1} (-1)^i \sum_{x=u_i}^{v_i} p(x, \mu_q)A_x - \sum_{i=2k}^{2m-1} (-1)^i \sum_{x=u_i}^{v_i} p(x, \mu_q)A_x - \sum_{x=u_{2m}}^{n} p(x, \mu_q)A_x.$$

Let

$$b_i(\mu_q) = \sum_{x=u_i}^{v_i} p(x, \mu_q)A_x$$
$$B_{i,j}(\mu_q) = \sum_{t=2i}^{2(i+j)-1} (-1)^t b_t(\mu_q)$$

then, since $A_x > 0$ for all x,

$$r(\mu_q, \delta) - r(\mu_q, \delta') \leqq B_{0,k}(\mu_q) - B_{k,m-k}(\mu_q).$$

A contradiction to (6) is then obtained by showing that there exists a pair (k, q) such that $2k - 1 \leqq q \leqq 2k$ and $B_{0,k}(\mu_q) < B_{k,m-k}(\mu_q)$.

Let $S_i(k_j)$ be the following statement, defined for all integer pairs (i, j) such that $0 \leqq i \leqq m - j$, $1 \leqq j \leqq m$.

$S_i(k_j)$: There exists an integer pair $(k_j, q(i, j))$ such that $0 \leqq k_j \leqq j$,

$$2(i + k_j) - 1 \leqq q(i, j) \leqq 2(i + k_j), \quad \text{and} \quad B_{i,k_j}(\mu_{q(i,j)}) < B_{i+k_j, j-k_j}(\mu_{q(i,j)}).$$

The negation of $S_i(k_j)$, written *not* $S_i(k_j)$, is then

not $S_i(k_j)$: For every integer pair (k, q) such that $0 \leqq k \leqq j$ and $2(i + k) - 1 \leqq q \leqq 2(i + k)$

$$B_{i,k}(\mu_q) \geqq B_{i+k,j-k}(\mu_q).$$

The desired contradiction to (6) may then be written $S_0(k_m)$.

The statement $S_i(k_1)$ is easily proved by contradiction. Note first that since δ belongs to C then

$$(7) \qquad \begin{aligned} p(u_{2i}, \mu_{2i}) &> 0 \qquad \text{for } i = 0, 1, \cdots, m - 1 \\ p(v_{2i-1}, \mu_{2i-1}) &> 0 \qquad \text{for } i = 1, 2, \cdots, m \end{aligned}$$

so that

$$(8) \qquad b_i(\mu_j) \binom{\geqq}{>} 0 \text{ for } i \binom{\neq}{=} j.$$

If $b_{2i+j}(\mu_{2i}) > 0$ then there exists $x_1 \geqq u_{2i+j}$ such that $p(x_1, \mu_{2i}) > 0$, and since $p(u_{2i}, \mu_{2i}) > 0$ then, by Lemma 1,

$$(9a) \qquad \text{if } b_{2i+j}(\mu_{2i}) > 0 \text{ then } p(x, \mu_{2i}) > 0 \text{ for } u_{2i} \leqq x \leqq u_{2i+j}.$$

Similarly, if $b_{2i}(\mu_{2i+j}) > 0$ then there exists $x_0 \leqq v_{2i}$ such that $p(x_0, \mu_{2i+j}) > 0$. By (7), however, there exists $x_1 \geqq u_{2i+j}$ such that $p(x_1, \mu_{2i+j}) > 0$; hence, by Lemma 1,

$$(9b) \qquad \text{if } b_{2i}(\mu_{2i+j}) > 0 \text{ then } p(x, \mu_{2i+j}) > 0 \text{ for } v_{2i} \leqq x \leqq u_{2i+j}.$$

It then follows that

$$(10) \qquad \text{if } B_{i,1}(\mu_{2i}) \leqq 0 \text{ then } B_{i,1}(\mu_{2i+j}) \binom{<}{\leqq} 0 \text{ for } j \binom{=}{>} 1$$

The statement in (10) is easily seen to hold for all $j \geqq 1$ such that either $b_{2i}(\mu_{2i+j}) = 0$ or $b_{2i+1}(\mu_{2i+j}) = 0$, for if $b_{2i}(\mu_{2i+j}) = 0$ then (8) implies (10) and if $b_{2i+1}(\mu_{2i+j}) = 0$ then $p(u_{2i+1}, \mu_{2i+j}) = 0$ but, by (7), there exists $x_1 > u_{2i+1}$ such that $p(x_1, \mu_{2i+j}) > 0$ so, by Lemma 1, $p(x, \mu_{2i+j}) = 0$ for all $x \leqq u_{2i+1}$ and, in particular, $b_{2i}(\mu_{2i+j}) = 0$. Now suppose that both $b_{2i}(\mu_{2i+j}) > 0$ and $b_{2i+1}(\mu_{2i+j}) > 0$ but that (10) does not hold; in particular, suppose $b_{2i+1}(\mu_{2i}) \geqq b_{2i}(\mu_{2i})$ and $b_{2i}(\mu_{2i+j}) \geqq b_{2i+1}(\mu_{2i+j})$. Since, by (8), $b_{2i}(\mu_{2i}) > 0$ then $b_{2i+1}(\mu_{2i}) > 0$ and, by (9a), $p(x, \mu_{2i}) > 0$ for $u_{2i} \leqq x \leqq u_{2i+1}$; and since $b_{2i}(\mu_{2i+j}) > 0$ then, by (9b), $p(x, \mu_{2i+j}) > 0$ for $v_{2i} \leqq x \leqq u_{2i+j}$. Then, by Condition 1,

$$\frac{b_{2i}(\mu_{2i})}{p(v_{2i}, \mu_{2i})} > \frac{b_{2i}(\mu_{2i+j})}{p(v_{2i}, \mu_{2i+j})};$$

but then the assumption that $b_{2i}(\mu_{2i+j}) \geqq b_{2i+1}(\mu_{2i+j})$ implies

$$\frac{a_{2i+1}(\mu_{2i})}{p(v_{2i}, \mu_{2i})} > \frac{b_{2i+1}(\mu_{2i+j})}{p(v_{2i}, \mu_{2i+j})},$$

which contradicts Condition 1. Hence (10) holds for all $j \geqq 1$. The statement

not $S_i(k_1)$ implies that $B_{i,1}(\mu_{2i}) \leqq 0$ and $B_{i,1}(\mu_{2i+1}) \geqq 0$ and is therefore a contradiction of (10); hence, $S_i(k_1)$ for all i such that $0 \leqq i \leqq m - 1$.

Now suppose $S_i(k_j)$ for all (i, j) such that $0 \leqq i \leqq m - j$, $1 \leqq j < s \leqq m$ but *not* $S_h(k_s)$, where $0 \leqq h \leqq m - s$. Then $(k_1, q(h, 1))$ can be chosen as $(0, 2h)$; otherwise $B_{h,1}(\mu_{2h}) \leqq 0$ and, since $2h < 2(h + 1 + k_{s-1}) - 1 \leqq q(h + 1, s - 1)$, then, by (10), $B_{h,1}(\mu_{q(h+1,s-1)}) \leqq 0$ which, together with the assumption $S_{h+1}(k_{s-1})$ implies the contradiction $S_h(k_s)$.

If for all $j < g < s$, $(k_j, q(h, j))$ can be chosen as $(0, 2h)$ then $(k_g, q(h, g))$ can be chosen as $(0, 2h)$. Otherwise, $B_{h,g}(\mu_{2h}) \leqq 0$ and, since $S_{h+g}(k_{s-g})$ but *not* $S_h(k_s)$, $B_{h,g}(\mu_{q(h+g,s-g)}) > 0$. And since $p(u_{2h}, \mu_{2h}) > 0$ then $p(x, \mu_{2h}) > 0$ for $u_{2h} \leqq x \leqq v_{2(h+g)-1}$; otherwise, by Lemma 1, $p(x, \mu_{2h}) = 0$ for all $x \geqq v_{2(h+g)-1}$, and then $(k_{g-1}, q(h, g - 1)) = (0, 2h)$ implies that $(k_g, q(h, g))$ can be chosen as $(0, 2h)$. Also, $p(v_{2(h+g-1)}, \mu_{q(h+g,s-g)}) > 0$; otherwise, by Lemma 1, $p(x, \mu_{q(h+g,s-g)}) = 0$ for all $x \leqq v_{2(h+g-1)}$ since $v_{2(h+g-1)} < v_{2(h+g+k_{s-g})-1} < u_{2(h+g+k_{s-g})}$, and then $B_{h,g}(\mu_{q(h+g,s-g)}) \leqq 0$ to contradict *not* $S_h(k_s)$. Hence,

$$-\frac{B_{h+g-1,1}(\mu_{2h})}{p(v_{2(h+g-1)}, \mu_{2h})} \geqq \frac{B_{h,g-1}(\mu_{2h})}{p(v_{2(h+g-1)}, \mu_{2h})} \tag{11}$$

and

$$\frac{B_{h,g-1}(\mu_{q(h+g,s-g)})}{p(v_{2(h+g-1)}, \mu_{q(h+g,s-g)})} \geqq -\frac{B_{h+g-1,1}(\mu_{q(h+g,s-g)})}{p(v_{2(h+g-1)}, \mu_{q(h+g,s-g)})}. \tag{12}$$

Observe, however, that if

$$\frac{B_{h,j}(\mu_{2h})}{p(v_{2(h+j)}, \mu_{2h})} > \frac{B_{h,j}(\mu_q)}{p(v_{2(h+j)}, \mu_q)} \tag{13}$$

where $1 \leqq j \leqq g - 1$, $2h < q$, and $p(x, \mu_q) > 0$ for $v_{2(h+j)} \leqq x \leqq v_{2(h+g)}$ then, by Condition 1,

$$\frac{B_{h+j,1}(\mu_{2h})}{p(v_{2(h+j)}, \mu_{2h})} > \frac{B_{h+j,1}(\mu_q)}{p(v_{2(h+j)}, \mu_q)} \tag{14}$$

and, since $B_{h,j}(\mu) + B_{h+j,1}(\mu) = B_{h,j+1}(\mu)$,

$$\frac{B_{h,j+1}(\mu_{2h})}{p(v_{2(h+j)}, \mu_{2h})} > \frac{B_{h,j+1}(\mu_q)}{p(v_{2(h+j)}, \mu_q)}. \tag{15}$$

Since $(k_{j+1}, q(h, j + 1))$ can be chosen as $(0, 2h)$ then $B_{h,j+1}(\mu_{2h}) > 0$, and since, by Condition 1, $p(v_{2(h+j)}, \mu_{2h})p(v_{2(h+j+1)}, \mu_q) > p(v_{2(h+j)}, \mu_q)p(v_{2(h+j+1)}, \mu_{2h})$ then

$$\begin{aligned} \frac{B_{h,j+1}(\mu_{2h})}{p(v_{2(h+j+1)}, \mu_{2h})} &= \frac{p(v_{2(h+j)}, \mu_{2h})}{p(v_{2(h+j+1)}, \mu_{2h})} \cdot \frac{B_{h,j+1}(\mu_{2h})}{p(v_{2(h+j)}, \mu_{2h})} \\ &> \frac{p(v_{2(h+j)}, \mu_q)}{p(v_{2(h+j+1)}, \mu_q)} \cdot \frac{B_{h,j+1}(\mu_q)}{p(v_{2(h+j)}, \mu_q)} = \frac{B_{h,j+1}(\mu_q)}{p(v_{2(h+j+1)}, \mu_q)}. \end{aligned} \tag{16}$$

Thus, if (13) then (16). Now let j' be the least j, $1 \leqq j \leqq g - 1$, such that $p(v_{2(h+j)}, \mu_{q(h+g,s-g)}) > 0$. Then (13) holds for $j = j'$ and $q = q(h + g, s - g)$,

for if $j' = 1$ and $p(v_{2h}, \mu_{q(h+g,s-g)}) > 0$, then let $j = 0$ in (14), (15), (16) to get the desired result; otherwise, if $j' \geqq 1$ and $p(v_{2h}, \mu_{q(h+g,s-g)}) = 0$ then the right side of (13) is nonpositive while the left side is positive since $(k_{j'}, q(h, j'))$ can be chosen as $(0, 2h)$. Hence, by finite induction, (13) holds for $j = g - 1$, $q = q(h + g, s - g)$; i.e.,

$$\frac{B_{h,g-1}(\mu_{2h})}{p(v_{2(h+g-1)}, \mu_{2h})} > \frac{B_{h,g-1}(\mu_{q(h+g,s-g)})}{p(v_{2(h+g-1)}, \mu_{q(h+g,s-g)})}.$$

Hence, by (11) and (12),

$$-\frac{B_{h+g-1,1}(\mu_{2h})}{p(v_{2(h+g-1)}, \mu_{2h})} > -\frac{B_{h+g-1,1}(\mu_{q(h+g,s-g)})}{p(v_{2(h+g-1)}, \mu_{q(h+g,s-g)})}$$

in contradiction to Condition 1. This proves that if $S_i(k_j)$ for all (i, j) such that $0 \leqq i \leqq m - j$, $1 \leqq j < s \leqq m$ but *not* $S_h(k_s)$ then $(k_j, q(h, j))$ can be chosen as $(0, 2h)$ for all j such that $1 \leqq j < s < m$.

With this result, however, simply take $j = s - 1$, $q = 2(h + s) - 1$ in (13) to get

$$\frac{B_{h,s-1}(\mu_{2h})}{p(v_{2(h+s-1)}, \mu_{2h})} > \frac{B_{h,s-1}(\mu_{2(h+s)-1})}{p(v_{2(h+s-1)}, \mu_{2(h+s)-1})}.$$

The denominator $p(v_{2(h+s-1)}, \mu_{2(h+s)-1})$ must be positive; otherwise, since $p(v_{2(h+s)-1}, \mu_{2(h+s)-1}) > 0$ then $p(x, \mu_{2(h+s)-1}) = 0$ for all $x \leqq v_{2(h+s-1)}$ so that $B_{h,s}(\mu_{2(h+s)-1}) = -b_{2(h+s)-1}(\mu_{2(h+s)-1}) < 0$ to contradict the assumption not $S_h(k_s)$. Then *not* $S_h(k_s)$ gives, as before,

$$-\frac{B_{h+s-1,1}(\mu_{2h})}{p(v_{2(h+s-1)}, \mu_{2h})} \geqq \frac{B_{h,s-1}(\mu_{2h})}{p(v_{2(h+s-1)}, \mu_{2h})}$$

$$\geqq \frac{B_{h,s-1}(\mu_{2(h+s)-1})}{p(v_{2(h+s-1)}, \mu_{2(h+s)-1})} \geqq \frac{B_{h+s-1,1}(\mu_{2(h+s)-1})}{p(v_{2(h+s-1)}, \mu_{2(h+s)-1})}$$

to contradict Condition 1. Hence, $S_h(k_s)$. And since $S_i(k_1)$ for all i such that $0 \leqq i \leqq m - 1$ then, by finite induction, $S_i(k_j)$ for all (i, j) such that $0 \leqq i \leqq m - j$, $1 \leqq j \leqq m$. In particular, $S_0(k_m)$, which establishes Theorem 3.

6. Minimax procedures when $W(\mu, \alpha) = |\alpha - \mu|$. When the [illegible] estimator does not have constant risk, as is obviously the [illegible] then the Bayes method of finding the minimax procedure [illegible] favorable a priori distribution becomes extremely difficult [illegible] distributions of the type considered here, however, [illegible] problem of guessing a least favorable a priori distribution [illegible] which points in the parameter space are assigned po[illegible] favorable a priori distribution.

THEOREM 4. *If $p(x, \mu)$ satisfies Conditions* 1, 2, [illegible] *then there exists a least favorable a priori distribution* [illegible] *ability to at most $n + 2$ values of μ, $\mu_0 \leqq \mu_1$* [illegible]

solution with respect to a least favorable priori distribution then $\mu_i \leqq \bar{\alpha}_{x_i} \leqq \mu_{i+1}$ *for* $i = 0, 1, \cdots, n$.

PROOF OF THEOREM 4. Assume, without loss of generality, that $M = (0, 1, \cdots, n)$. Let

$$r_x(\xi, \alpha) = \sum_{\mu=0}^{N} |\alpha - \mu| \, p(x, \mu)\xi_\mu$$

and let a_x^ξ be the collection of integers

$$a_x^\xi = (\alpha, 0 \leqq \alpha \leqq N \mid \inf_\beta r_x(\xi, \beta) = r_x(\xi, \alpha).$$

From the proof of Theorem 2 the function $r_x(\xi, \alpha)$ has the properties

I′: if $r_x(\xi, \alpha) \leqq r_x(\xi, \alpha + 1)$ then $r_x(\xi, \alpha + \beta) \leqq r_x(\xi, \alpha + \beta + 1)$ for all $\beta \geqq 0$

II: if $r_y(\xi, \alpha) \leqq r_y(\xi, \alpha + 1)$ then $r_x(\xi, \alpha) < r_x(\xi, \alpha + 1)$ for all $x \leqq y$

Hence, a_x^ξ has the form

$$a_x^\xi = (\alpha_x^\xi, \alpha_x^\xi + 1, \cdots, \alpha_x^\xi + \beta_x^\xi)$$

where $0 \leqq \alpha_x^\xi \leqq N$, $0 \leqq \beta_x^\xi \leqq N - \alpha_x^\xi$, and $\alpha_x^\xi + \beta_x^\xi \leqq \alpha_{x+1}^\xi$. Furthermore, since $r_x(\xi, \alpha_x^\xi) = r_x(\xi, \alpha_x^\xi + 1) = \cdots = r_x(\xi, \alpha_x^\xi + \beta_x^\xi)$ or, for $i = 1, \cdots, \beta_x^\xi - 1$,

$$p(x, \alpha_x^\xi + i)\xi_{\alpha_x^\xi+i} + \sum_{\mu=\alpha_x^\xi+i+1}^{N} p(x, \mu)\xi_\mu = \sum_{\mu=0}^{\alpha_x^\xi+i-1} p(x, \mu)\xi_\mu$$

and

$$\sum_{\mu=\alpha_x^\xi+i+1}^{N} p(x, \mu)\xi_\mu = \sum_{\mu=0}^{\alpha_x^\xi+i-1} p(x, \mu)\xi_\mu + p(x, \alpha_x^\xi + i)\xi_{\alpha_x^\xi+i}$$

then $p(x, \alpha_x^\xi + i)\xi_{\alpha_x^\xi+i} = 0$ for $0 < i < \beta_x$. Hence, since $r_x(\xi, \alpha_x^\xi - 1) > r_x(\xi, \alpha_x^\xi)$, or

$$p(x, \alpha_x^\xi)\xi_{\alpha_x^\xi} + p(x, \alpha_x^\xi + \beta_x^\xi)\xi_{\alpha_x^\xi+\beta_x^\xi} + \sum_{\mu=\alpha_x^\xi+\beta_x^\xi-1}^{N} p(x, \mu)\xi_\mu > \sum_{u=0}^{\alpha_x^\xi-1} p\,(x, \mu)\xi_\mu$$

and $r_x(\xi, \alpha_x^\xi + \beta_x^\xi) < r_x(\xi, \alpha_x^\xi + \beta_x^\xi + 1)$, or

$$\sum_{\mu=\alpha_x^\xi+\beta_x^\xi+1}^{N} p(x, \mu)\xi_\mu < \sum_{\mu=0}^{\alpha_x^\xi-1} p(x, \mu)\xi_\mu + p(x, \alpha_x^\xi)\xi_{\alpha_x^\xi} + p(x, \alpha_x^\xi + \beta_x^\xi)\xi_{\alpha_x^\xi+\beta_x^\xi},$$

then $p(x, \alpha_x^\xi)\xi_{\alpha_x^\xi} > 0$ and $p(x, \alpha_x^\xi + \beta_x^\xi)\xi_{\alpha_x^\xi+\beta_x^\xi} > 0$. Therefore, since $p(x, \alpha_x^\xi) > 0$ and $p(x, \alpha_x^\xi + \beta_x^\xi) > 0$ imply, by Lemma 1, that $p(x, \alpha_x^\xi + i) > 0$ for $0 \leqq i \leqq \beta_x^\xi$, then

$$\text{(17)} \qquad \xi_{\alpha_x^\xi+i} \begin{cases} > 0 \text{ for } i = 0 \\ = 0 \text{ for } 0 < i < \beta_x^\xi \\ > 0 \text{ for } i = \beta_x^\xi . \end{cases}$$

If ξ^0 is a least favorable a priori distribution; i.e., if ξ^0 maximizes $\inf_\delta r(\xi, \delta)$,

then since, for a fixed δ, $r(\xi, \delta)$ is linear in ξ, every ξ which satisfies

(18) $$r_x(\xi, \alpha_x^{\xi^0} - 1) \geqq r_x(\xi, \alpha_x^{\xi^0}) = \cdots = r_x(\xi, \alpha_x^{\xi^0} + \beta_x^{\xi^0}) \leqq r_x(\xi, \alpha_x^{\xi^0} + \beta_x^{\xi^0} + 1)$$

for $x = 0, 1, \cdots, n$ is likewise a least favorable a priori distribution. But since $p(x, \mu)$ satisfies Conditions 1, 2, and 3, and, for every $x > 0$, $p(x - 1, \alpha_{x-1}^{\xi^0} + \beta_{x-1}^{\xi^0}) > 0$ and $p(x, \alpha_x^{\xi^0}) > 0$, where $\alpha_{x-1}^{\xi^0} + \beta_{x-1}^{\xi^0} \leqq \alpha_x^{\xi^0}$, then for every $x > 0$ there exists an integer μ_x such that $\alpha_{x-1}^{\xi^0} + \beta_{x-1}^{\xi^0} \leqq \mu_x \leqq \alpha_x^{\xi^0}$ and both $p(x - 1, \mu_x) > 0$ and $p(x, \mu_x) > 0$. Let (μ_x), $x = 1, 2, \cdots, n$, $\mu_x \leqq \mu_{x+1}$, be a sequence of such integers and define $\mu_0 = \alpha_0^{\xi^0}$ and $\mu_{n+1} = \alpha_n^{\xi^0} + \beta_n^{\xi^0}$. Then every ξ which satisfies

(19) $$r_x(\xi, \mu_x) = r_x(\xi, \mu_x + 1) = \cdots = r_x(\xi, \mu_{x+1})$$

for $x = 0, 1, \cdots, n$ also satisfies (18) and has $\xi_\mu = 0$ for $\mu_x < \mu < \mu_{x+1}$ for $x = 0, 1, \cdots, n$.

It remains, then, to show that a solution ξ' to (19) exists and may be chosen so that $\xi'_\mu = 0$ for $\mu < \mu_0$ and for $\mu > \mu_{n+1}$. This, however, follows directly from Theorem 3, for the problem of proving the existence of such a ξ' is easily seen to reduce to the problem of proving that a set of equations of the form

$$\sum_{\mu=0}^{x} p'(x, \mu)\xi_\mu = \sum_{\mu=x+1}^{m} p'(x, \mu)\xi_\mu, x = 0, 1, \cdots, m - 1$$

where $p'(x, \mu)$ satisfies Conditions 1 and 2 for $x = 0, 1, \cdots, n \geqq m - 1$, $\mu = 0, 1, \cdots, m$, and $p'(x, x) > 0$ and $p'(x, x + 1) > 0$, has a solution $\xi = (\xi_\mu)$ such that $\xi_\mu > 0$, $\mu = 0, 1, \cdots, m$, and $\sum_{\mu=0}^{m} \xi_\mu = 1$, and this may be viewed as a special case of Theorem 3 with $N = m$ and $n \geqq m - 1$. Theorem 3 then asserts that a procedure δ with $\delta_x(x) + \delta_{x+1}(x) = 1$, $\delta_x(x) < 1$, for $x = 0, 1, \cdots, m - 1$ and $\delta_m(x) = 1$ for $x \geqq m$ is admissible, and therefore δ is a Bayes solution relative to some a priori distribution ξ and, by (17), $\xi_\mu > 0$ for $\mu = 0, 1, \cdots, m$. Hence, a ξ' of the desired form exists and the theorem is established.

The construction of the minimax procedure δ^0 is easily accomplished once the integer μ_x is known for every x. δ^0 is defined by $(\bar{\alpha}_0^0, \bar{\alpha}_1^0, \cdots, \bar{\alpha}_n^0)$ which is uniquely determined by the equations

$$r(\mu_x, \delta^0) = \sum_{y=0}^{x-1} p(y, \mu_x)(\mu_x - \bar{\alpha}_y^0) + \sum_{y=x}^{n} p(y, \mu_x)(\bar{\alpha}_y^0 - \mu) = r(\mu_0, \delta^0)$$

7. Discussion. The requirement that an estimator of an integer-valued parameter must itself be integer-valued is almost a logical necessity in any rigorous approach to the estimation problem. For practical purposes, of course, such a requirement has been regarded as an unnecessary refinement, and statisticians conventionally estimate an integer-valued parameter by means of a real-valued statistic, presenting as their estimate either the real number itself or the nearest integer. The problem is frequently encountered, for example, in such a form that the statistician wishes to present an estimate of the fraction μ/N. Certainly, division by the known constant N is a trivial alteration of the estima-

tion problem; it would be unheard of, however, to require in this case that the estimate assume one of the values $0/N$, $1/N$, $\cdots$, N/N.

If real-valued procedures are allowed then when loss is absolute error the randomized, integer-valued procedure δ is equivalent to the non-randomized procedure which estimates the real number $\bar{a}_x$ when x is observed. Any optimum property ascribed to an integer-valued procedure therefore applies to its real-valued counterpart so, as a corollary to Theorem 3, when real-valued procedures are allowed then the class of non-randomized real-valued procedures derived from the class C in the above manner is a minimal essentially complete class. Likewise, if δ^0 is the minimax integer-valued procedure then the non-randomized real-valued procedure $\bar{a}_x^0$ is also minimax. Theorems 3 and 4 thus remain essentially unaffected by the introduction of real-valued procedures.

REFERENCES

[1] H. Rubin, "A complete class of decision procedures for distributions with monotone likelihood ratio," Technical Report 9, Dept. of Stat., Stanford University, October 31, 1951.

[2] S. Karlin and H. Rubin, "The theory of decision procedure for distributions with monotone likelihood ratio." *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272–299.

MAXIMUM-LIKELIHOOD ESTIMATION OF PARAMETERS SUBJECT TO RESTRAINTS

By J. Aitchison and S. D. Silvey

University of Glasgow

Summary. The estimation of a parameter lying in a subset of a set of possible parameters is considered. This subset is the null space of a well-behaved function and the estimator considered lies in the subset and is a solution of likelihood equations containing a Lagrangian multiplier. It is proved that, under certain conditions analogous to those of Cramér, these equations have a solution which gives a local maximum of the likelihood function. The asymptotic distribution of this 'restricted maximum likelihood estimator' and an iterative method of solving the equations are discussed. Finally a test is introduced of the hypothesis that the true parameter does lie in the subset; this test, which is of wide applicability, makes use of the distribution of the random Lagrangian multiplier appearing in the likelihood equations.

1. Introduction. Quite frequently in statistical theory the natural way of building up a mathematical model of an experiment leads to the description of the experiment by a random variable X whose distribution function F depends on s parameters $\theta_1, \theta_2, \cdots, \theta_s$, which are not mathematically independent but satisfy r functional relationships $h_i(\theta_1, \theta_2, \cdots, \theta_s) = 0$, $i = 1, 2, \cdots, r$, $r < s$. In many cases where such a natural description arises it is possible to solve the r equations $h_i(\theta_1, \theta_2, \cdots, \theta_s) = 0$ for r of the parameters in terms of the remaining $s - r$, to express the distribution function F in terms of these remaining parameters only and, given observations on X, to estimate these $s - r$ unrestricted parameters by the method of maximum likelihood. This procedure has two disadvantages. First, it may be impossible to express r of the parameters explicitly in terms of the remaining $s - r$ and second, interest may lie in estimating all of the parameters simultaneously, in which case a symmetrical procedure for so doing is certainly desirable. The natural symmetric method for maximum-likelihood estimation in this case is achieved by the introduction of Lagrangian multipliers and it is this method that we will consider in this paper.

2. Formulation of the problem. In this section we will formulate more precisely the problem to be considered.

We will denote m-dimensional Euclidian space by $\mathfrak{R}^m$, $m = 1, 2, 3, \cdots$. A point in $\mathfrak{R}^s$, denoted by $\theta = (\theta_1, \theta_2, \cdots, \theta_s)$ will represent a value of a parameter. There is a particular point $\theta_0 = (\theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_s^{(0)})$ in $\mathfrak{R}^s$ which is the true, though unknown, parameter value. Corresponding to each θ in some neighbour-

Received February 18, 1957.

hood of θ_0, say in $U_\alpha = \{\theta: \|\theta - \theta_0\| \leqq \alpha\}$, is a probability density function f_θ defined on $\mathcal{R}^1$ and we will denote the value of f_θ at the point $t \in \mathcal{R}^1$ by $f(t, \theta)$. The probability density function f_{θ_0} defines a probability measure on $\mathcal{R}^1$ and we will assume that, with respect to this measure, for almost all t, the partial derivatives $\partial \log f(t, \theta)/\partial\theta_i$, $i = 1, 2, \cdots, s$, exist for every θ in U_α.

There is given a continuous function h from $\mathcal{R}^s$ into $\mathcal{R}^r$, $r < s$, defined by $h(\theta) = (h_1(\theta), h_2(\theta), \cdots, h_r(\theta))$, which is such that, for every θ in U_α, the partial derivatives $\partial h_j(\theta)/\partial\theta_i$, $i = 1, 2, \cdots, s, j = 1, 2, \cdots, r$, exist. The function h has the further property that $h(\theta_0) = 0$.

A point in $\mathcal{R}^n$ denoted by $x = (x_1, x_2, \cdots, x_n)$ will be regarded as representing a set of n independent observations on a random variable whose probability density function is f_{θ_0} and we use the fact that points in $\mathcal{R}^n$ are being so regarded to define, in the usual way, a probability measure on $\mathcal{R}^n$, for each n. Subsequent statements regarding the probabilities of sets in $\mathcal{R}^n$ will refer to this particular probability measure.

It will be convenient to use also matrix representation for points in $\mathcal{R}^m$ and for linear operators from one Euclidian space to another and we will use the convention that, for example, $\boldsymbol{\theta}$ is the $s \times 1$ column vector representing the point θ in $\mathcal{R}^s$, and $\mathbf{H}$, an $s \times r$ matrix, represents a linear operator H from $\mathcal{R}^r$ into $\mathcal{R}^s$.

The log-likelihood function L is defined on a subset of $\mathcal{R}^n \times \mathcal{R}^s$ by

$$L(x, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta).$$

If $\mathbf{H}_\theta$ denotes the $s \times r$ matrix $(\partial h_j(\theta)/\partial\theta_i)$, and if λ is a Lagrangian multiplier in $\mathcal{R}^r$, then we propose to estimate the unknown parameter θ_0 by a solution, if such exists, of the equations

$$\ell(x, \theta) + H_\theta\lambda = 0 \tag{2.1}$$

$$h(\theta) = 0, \tag{2.2}$$

where $\ell(x, \theta)$ is the point in $\mathcal{R}^s$ whose ith component is $\partial L(x, \theta)/\partial\theta_i$.

We will show that, under certain fairly general conditions, if x belongs to a set whose probability measure tends to 1 as $n \to \infty$, these equations have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$, where $\hat{\theta}(x)$ is near θ_0 and $\hat{\theta}(x)$ maximises $L(x, \theta)$ subject to the condition $h(\theta) = 0$. The definition of $\hat{\theta}$ and $\hat{\lambda}$ will then be extended in a natural way to the whole of $\mathcal{R}^n$ and we will show that the random variables thus defined are asymptotically jointly normally distributed. We will then consider an iterative procedure for solving the equations (2.1) and (2.2). Finally tests of the adequacy of the model will be introduced.

3. Existence of a solution. The proof that we will give of the existence of a solution of the equations (2.1) and (2.2) is based on the same principle as a proof given by Cramér [2] of the existence of a maximum likelihood estimate of a parameter in $\mathcal{R}^1$. However the presence of the restraining condition $h(\theta) = 0$ in the situation we are discussing makes our proof more intricate in detail than a

straightforward generalisation of Cramér's proof to a parameter in $\mathcal{R}^s$ would be. And we start by indicating the main lines of the proof.

We set out to show that, under certain conditions, if δ is a sufficiently small given number and if n is sufficiently large, then, for a set of x whose probability measure is near 1, the equations (2.1) and (2.2) have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$, where $\hat{\theta}(x) \varepsilon U_\delta$. We will demand that in U_α the function $\log f(x, \cdot)$ should possess partial derivatives of the third order and the components of the function h should possess partial derivatives of the second order. Then it will be possible, by expanding the components of $\ell(x, \theta)$ and $h(\theta)$ about θ_0 to express the equations (in matrix notation) in the form

$$\text{(3.1)} \qquad \mathbf{l}(x, \theta_0) + \mathbf{M}_{x, \theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \mathbf{v}^{(1)}(x, \theta) + \mathbf{H}_\theta \boldsymbol{\lambda} = \mathbf{0},$$

$$\text{(3.2)} \qquad \mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \mathbf{v}^{(2)}(\theta) = \mathbf{0},$$

where

(i) $\mathbf{M}_{x, \theta_0}$ is the matrix $(\partial^2 L(x, \theta_0)/\partial\theta_i\partial\theta_j)$,

(ii) $\mathbf{v}^{(1)}(x, \theta)$ is a vector whose mth component may be expressed in the form $\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{L}_m(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, $\mathbf{L}_m$ being the matrix $(\partial^3 L(x, \theta^{(m,1)})/\partial\theta_m\partial\theta_i\partial\theta_j)$, $i, j = 1, 2, \cdots, s$, and $\theta^{(m,1)}$ a point such that $\| \theta^{(m,1)} - \theta_0 \| < \| \theta - \theta_0 \|$.

(iii) $\mathbf{v}^{(2)}(\theta)$ is a vector whose mth component is $\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{H}_m(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, $\mathbf{H}_m$ being the matrix $(\partial^2 h_m(\theta^{(m,2)})/\partial\theta_i\partial\theta_j)$, $i, j = 1, 2, \cdots, s$, and $\theta^{(m,2)}$ a point such that $\| \theta^{(m,2)} - \theta_0 \| < \| \theta - \theta_0 \|$.

Further conditions imposed on f, which are almost a straightforward generalisation of Cramér's conditions [2], will ensure that, for large enough n, there is a set of x whose probability measure is near 1 such that, if x belongs to this set,

(i) $\| (1/n)\ell(x, \theta_0) \|$ is small,

(ii) $-(1/n)\mathbf{M}_{x, \theta_0}$ is near a certain positive definite matrix $\mathbf{B}_{\theta_0}$ and

(iii) the elements of $(1/n)\mathbf{L}_m$ are bounded for $\theta \varepsilon U_\delta$. By dividing (3.1) by n we will then be able to express this equation in the form

$$\text{(3.3)} \qquad -\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{n}\mathbf{H}_\theta\boldsymbol{\lambda} + \delta^2\mathbf{v}^{(3)}(x, \theta) = \mathbf{0}$$

where $\| v^{(3)}(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$. In addition we will demand that, for $\theta \varepsilon U_\alpha$, the second order derivatives of the components of h should be bounded. Then we will be able to express (3.2) in the form

$$\text{(3.4)} \qquad \mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \delta^2\mathbf{v}^{(4)}(\theta) = \mathbf{0}$$

where $\| v^{(4)}(\theta) \|$ is bounded for $\theta \varepsilon U_\delta$.

If the equations (3.3) and (3.4) have a solution, then by pre-multiplying (3.3) by $\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}$ and substituting for $\mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ from (3.4) we find that the values of θ and λ satisfying these equations also satisfy an equation of the form

$$\text{(3.5)} \qquad \mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}\mathbf{H}_\theta\left(\frac{1}{n}\boldsymbol{\lambda}\right) + \delta^2\mathbf{v}^{(5)}(x, \theta) = \mathbf{0}.$$

We will impose conditions on h which ensure that the matrix $\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}\mathbf{H}_\theta$ is non-

singular and the elements of its inverse are bounded functions of θ for $\theta \varepsilon U_\delta$. Then it will be possible to solve equation (3.5) for λ in terms of θ and on substitution in (3.3) we will obtain the result that any value of θ in U_δ for which equations (3.3) and (3.4) are satisfied is also a solution of an equation of the form

$$(3.6) \qquad -\mathbf{B}_{\theta_0}(\theta - \theta_0) + \delta^2 \mathbf{v}(x, \theta) = 0$$

where $\| v(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$.

Conversely it will be shown that if the equation (3.6) has a solution $\hat{\theta}(x) \varepsilon U_\delta$ then $\hat{\theta}(x)$ leads to a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$ of equations (2.1) and (2.2). We will then use the fact that $\mathbf{B}_{\theta_0}$ is a positive definite matrix to prove that, if δ is sufficiently small, (3.6) has a solution in U_δ.

This outline of the method of proof to be adopted provides the motivation for the introduction of conditions on f and h which we now discuss.

Conditions on f. The following conditions on the function f appear complicated and restrictive from the mathematical point of view. In fact they will be satisfied in most practical estimation problems.

$\mathfrak{F}1$. *For every $\theta \varepsilon U_\alpha$ and for almost all $t \varepsilon \mathcal{R}^1$ (almost all with respect to the probability measure on $\mathcal{R}^1$ defined by f_{θ_0}), the derivatives*

$$\frac{\partial \log f(t, \theta)}{\partial \theta_i}, \quad \frac{\partial^2 \log f(t, \theta)}{\partial \theta_i \partial \theta_j} \quad \text{and} \quad \frac{\partial^3 \log f(t, \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}, \qquad i, j, k = 1, 2, \cdots, s,$$

exist, and the first and second order derivatives are continuous functions of θ.

$\mathfrak{F}2$. *For every $\theta \varepsilon U_\alpha$ and for $i, j = 1, 2, \cdots, s$, $| \partial f(t, \theta)/\partial\theta_i | < F_1(t)$ and $| \partial^2 f(t, \theta)/\partial\theta_i\partial\theta_j | < F_2(t)$, where F_1 and F_2 are finitely integrable over $(-\infty, \infty)$.*

$\mathfrak{F}3$. *For every $\theta \varepsilon U_\alpha$ and $i, j, k = 1, 2, \cdots, s$, $| \partial^3 \log f(t, \theta)/\partial\theta_i\partial\theta_j\partial\theta_k | < F_3(t)$, where $\int_{-\infty}^{\infty} F_3(t) f(t, \theta_0)\, dt$ is finite and equal to κ_1, say.*

$$\mathfrak{F}4. \qquad b_{ij} = \int_{-\infty}^{\infty} \frac{\partial \log f(t, \theta_0)}{\partial \theta_i} \frac{\partial \log f(t, \theta_0)}{\partial \theta_j} f(t, \theta_0)\, dt$$

is finite for $i, j = 1, 2, \cdots, s$, and the matrix $\mathbf{B}_{\theta_0} = (b_{ij})$ is positive definite with minimum latent root μ_0.

The conditions $\mathfrak{F}3$ and $\mathfrak{F}4$ are apparently less stringent than a straightforward generalisation of Cramér's corresponding conditions would be. In §6 we return to this point.

If f satisfies these conditions then for any given positive numbers $\delta < \alpha$ and $\epsilon < 1$ and for sufficiently large n, say $n \geqq n(\delta, \epsilon)$, there exists a set $X_n \subset \mathcal{R}^n$ with the properties

$\mathfrak{X}1$. $\Pr\{X_n\} > 1 - \epsilon$.

$\mathfrak{X}2$. $\left\| \frac{1}{n} \ell(x, \theta_0) \right\| < \delta^2$, if $x \varepsilon X_n$.

$\mathfrak{X}3$. $\frac{1}{n} \mathbf{M}_{x, \theta_0}$ can be expressed in the form $-\mathbf{B}_{\theta_0} + \delta \mathbf{m}_{x,\theta_0}$,

where $\mathbf{m}_{x,\theta_0}$ is an $s \times s$ matrix the moduli of whose elements are bounded by 1, if $x \varepsilon X_n$.

$\mathfrak{X}4$. For *every* $\theta \varepsilon U_\alpha$ and $i, j, k = 1, 2, \cdots, s$,

$$\left| \frac{1}{n} \frac{\partial^3 L(x, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < 2\kappa_1$$

if $x \varepsilon X_n$.

The proof of these results is similar to the proof of the corresponding results given by Cramér [2] in the case of a parameter in $\mathfrak{R}^1$ and we merely remark that the conditions $\mathfrak{F}1$–4 imply (as they are designed to imply) that

(i) $(1/n)\ell(\cdot, \theta_0)$ converges in probability to $0 \varepsilon \mathfrak{R}^s$,

(ii) $(1/n)\mathbf{M}_{\cdot,\theta_0}$ converges in probability to $-\mathbf{B}_{\theta_0}$, and

(iii) if $G(x) = 1/n \sum_{i=1}^{n} F_3(x_i)$, then the random variable G converges in probability to κ_1 and

$$\frac{1}{n} \left| \frac{\partial^3 L(x, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| = \frac{1}{n} \left| \sum_{l=1}^{n} \frac{\partial^3 \log f(x_l, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < G(x),$$

by $\mathfrak{F}3$.

In future when we refer to a set X_n we imply that n is sufficiently large for the existence of a set in $\mathfrak{R}^n$ with the properties $\mathfrak{X}1$–4 and that the set X_n referred to has these properties.

As has already been indicated, one of the main purposes of the introduction of the conditions $\mathfrak{F}$ was to ensure that (3.1) could be expressed in the form (3.3). Now if the conditions $\mathfrak{F}$ are satisfied, if $x \varepsilon X_n$ and $\theta \varepsilon U_\delta$, it is easily verified that

(i) by $\mathfrak{X}2$, $(1/n\delta^2)\| \ell(x, \theta_0) \| < 1$,

(ii) by $\mathfrak{X}3$, $(1/\delta)\| m_{x,\theta_0}(\theta - \theta_0) \| \leqq s^2$,

(iii) by $\mathfrak{X}4$, $(1/n\delta^2)\| v^{(1)}(x, \theta) \| < (1/\delta^2)s^3\kappa_1\| \theta - \theta_0 \|^2 \leqq s^3\kappa_1$.

It follows that (3.1) can then be expressed in the form (3.3) and

$$\| v^{(3)}(x, \theta) \| < 1 + s^2 + s^3\kappa_1, \quad \text{when} \quad x \varepsilon X_n \quad \text{and} \quad \theta \varepsilon U_\delta.$$

Conditions on h. We impose the following conditions on the function h.

$\mathfrak{H}1$. *For every $\theta \varepsilon U_\alpha$ the partial derivatives $\partial h_k(\theta)/\partial\theta_i$, $i = 1, 2, \cdots, s$, $k = 1, 2, \cdots, r$, exist and these are continuous functions of θ.*

$\mathfrak{H}2$. *For every $\theta \varepsilon U_\alpha$ the partial derivatives $\partial^2 h_k(\theta)/\partial\theta_i\partial\theta_j$, $i, j = 1, 2, \cdots, s$, $k = 1, 2, \cdots, r$, exist and $| \partial^2 h_k(\theta)/\partial\theta_i\partial\theta_j | < 2\kappa_2$, a given constant, for all i, j and k.*

$\mathfrak{H}3$. *The $s \times r$ matrix $\mathbf{H}_{\theta_0}$ is of rank r.*

The condition $\mathfrak{H}2$ is introduced to ensure that when (3.2) is expressed in the form (3.4), $\| v^{(4)}(\theta) \|$ is bounded for $\theta \varepsilon U_\delta$. It is clear that it does ensure this since, as is easily verified, by $\mathfrak{H}2$, $\| v^{(2)}(\theta) \| < s^2\kappa_2 \| \theta - \theta_0 \|^2$ and so $\| v^{(4)}(\theta) \| = (1/\delta^2)\| v^{(2)}(\theta) \| < s^2\kappa_2$ if $\theta \varepsilon U_\delta$.

Also the condition $\mathcal{H}3$ implies that the matrix $\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_{\theta_0}$ is positive definite since the matrix $\mathbf{B}_{\theta_0}^{-1}$ is positive definite. Since the elements of $\mathbf{H}_\theta$ are, by $\mathcal{H}1$ continuous functions of θ it follows that there exists a neighbourhood of θ in which det $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)$ is bounded away from zero, and we may assume that this neighbourhood is U_α. (This assumption merely involves choosing α small enough initially). This means that when $\theta \varepsilon U_\alpha$ we can solve the equation (3.5) for λ in terms of θ. Furthermore the elements of the matrix $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}$ are then continuous functions on U_α since the elements of $\mathbf{H}_\theta$ are continuous and det $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)$ is bounded away from zero. Since U_α is a closed set it follows that the elements of $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}$ are uniformly bounded on U_α. This result, together with the results that $\| v^{(3)}(x, \theta) \|$ and $\| v^{(4)}(\theta) \|$ are bounded on U_δ, enable us to prove that when λ is eliminated from (3.3) and (3.4), and (3.6) is obtained, then in (3.6) $\| v(x, \theta) \|$ is bounded on U_δ, if $x \varepsilon X_n$.

We have now gone a considerable way towards proving the main part of the following lemma.

LEMMA 1. *Subject to the conditions* $\mathcal{F}$ *and* $\mathcal{H}$, *if* $\delta < \alpha$ *and* $\epsilon < 1$ *are given positive numbers and if* $x \varepsilon X_n$, *then the equations* (2.1) *and* (2.2) *have a solution* $\hat\theta(x)$, $\hat\lambda(x)$ *such that* $\hat\theta(x) \varepsilon U_\delta$, *if and only if* $\hat\theta(x)$ *satisfies a certain equation of the form* $-B_{\theta_0}(\theta - \theta_0) + \delta^2 v(x, \theta) = 0$. *In this equation* $v(x, \cdot)$ *is a continuous function on* U_δ *and* $\| v(x, \theta) \|$ *is bounded for* $\theta \varepsilon U_\delta$ *by a positive number* κ_3, *say.*

PROOF. The fact that the condition is necessary has virtually been established already. On eliminating λ from (2.1) and (2.2) by the method outlined at the beginning of §3 we obtain, in matrix notation, the following explicit expression for (3.6)

$$-\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \mathbf{H}_\theta(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}\{\mathbf{v}^{(2)}(\theta) + \mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{v}^{(6)}(x, \theta)\} + \mathbf{v}^{(6)}(x, \theta) = 0, \tag{3.7}$$

where

$$\mathbf{v}^{(6)}(x, \theta) = \delta^2\mathbf{v}^{(3)}(x, \theta) = \frac{1}{n}\mathbf{l}(x, \theta) + \mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \tag{3.8}$$

and

$$\mathbf{v}^{(2)}(\theta) = \delta^2\mathbf{v}^{(4)}(\theta) = \mathbf{h}(\theta) - \mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0). \tag{3.9}$$

Hence in (3.6),

$$\mathbf{v}(x, \theta) = -\mathbf{H}_\theta(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}\{\mathbf{v}^{(4)}(\theta) + \mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{v}^{(3)}(x, \theta)\} + \mathbf{v}^{(3)}(x, \theta). \tag{3.10}$$

The fact that $v(x, \cdot)$ is a continuous function on U_δ and that $\| v(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$ follows from (3.8), (3.9) and (3.10), in virtue of the discussion of $v^{(3)}(x, \theta)$, $v^{(4)}(\theta)$ and $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}$ above.

Turning to the sufficiency of the condition we now suppose that the equation

(3.7) has a root $\hat{\theta}(x) \in U_\delta$. Then, writing $\hat{\theta}$ instead of $\hat{\theta}(x)$ for brevity, we obtain on premultiplication of (3.7) by $\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}$,

$$-\mathbf{H}'_{\theta_0}(\hat{\theta} - \theta_0) - \mathbf{v}^{(2)}(\hat{\theta}) = 0, \tag{3.11}$$

i.e., by (3.9),

$$\mathbf{h}(\hat{\theta}) = 0.$$

Substitution for $\mathbf{v}^{(2)}(\hat{\theta})$ from (3.11) and for $\mathbf{v}^{(6)}(x, \hat{\theta})$ from (3.8), in (3.7) gives

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{H}_{\hat{\theta}}(\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}\mathbf{H}_{\hat{\theta}})^{-1}\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}\mathbf{l}(x, \hat{\theta}),$$

or, if we write $\mathbf{Q}_{\hat{\theta}}$ for $(\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}\mathbf{H}_{\hat{\theta}})^{-1}\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}$,

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{H}_{\hat{\theta}}\mathbf{Q}_{\hat{\theta}}\mathbf{l}(x, \hat{\theta}). \tag{3.12}$$

If we now define $\hat{\lambda}(x)$ by

$$\hat{\lambda}(x) = -\mathbf{Q}_{\hat{\theta}}\mathbf{l}(x, \hat{\theta}),$$

then

$$\mathbf{l}(x, \hat{\theta}) + \mathbf{H}_{\hat{\theta}}\hat{\lambda}(x) = 0,$$

and $\hat{\theta}(x)$, $\hat{\lambda}(x)$ satisfy the equations (2.1) and (2.2)

In order to prove that the equation (3.6) has a root in U_δ, if δ is sufficiently small, we will require the following lemma.

LEMMA 2. *If g is a continuous function mapping $\mathcal{R}^s$ into itself with the property that, for every θ such that $\| \theta \| = 1$, $\theta' g(\theta) < 0$, then there exists a point $\hat{\theta}$ such that $\| \hat{\theta} \| < 1$* and *$g(\hat{\theta}) = 0$.*

PROOF. For the proof of this result we are indebted to Mr. J. M. Michael who has proved that this result is equivalent to Brouwer's fixed point theorem [4]. A direct proof from the latter theorem is as follows.

We suppose that $g(\theta) \neq 0$ for any θ such that $\| \theta \| \leq 1$ Then the function g_1, defined on the unit sphere in $\mathcal{R}^s$ by

$$g_1(\theta) = \frac{g(\theta)}{\| g(\theta) \|},$$

is a continuous function mapping this unit sphere into itself. Hence by the fixed point theorem there is a point θ^* in the unit sphere such that $\theta^* = g_1(\theta^*)$. Also since $\| g_1(\theta) \| = 1$ for every θ in the unit sphere, it follows that $\| \theta^* \| = 1$, and $\theta^{*\prime} g_1(\theta^*) = \theta^{*\prime}\theta^* = 1 > 0$. But this contradicts the fact that $\theta' g(\theta) < 0$ (and consequently that $\theta' g_1(\theta) < 0$) for every θ such that $\| \theta \| = 1$.

Hence there is a point $\hat{\theta}$ in the unit sphere such that $g(\hat{\theta}) = 0$. It is obvious that $\| \hat{\theta} \| \neq 1$. Hence $\| \hat{\theta} \| < 1$.

We are now in a position to prove the following existence theorem.

THEOREM 1. *Subject to the conditions $\mathfrak{F}$ and $\mathfrak{H}$, if δ is a sufficiently small given*

positive number, ϵ is a given positive number less than 1 and if $x \varepsilon X_n$, then the equations (2.1) and (2.2) have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$ such that $\hat{\theta}(x) \varepsilon U_\delta$.

PROOF. We suppose $\delta < \alpha$ and $x \varepsilon X_n$. We consider (3.6) and define a function g on the unit sphere in $\mathcal{R}^s$ by

$$g\left(\frac{\theta - \theta_0}{\delta}\right) = -B_{\theta_0}(\theta - \theta_0) + \delta^2 v(x, \theta).$$

By Lemma 1, $v(x, \cdot)$ is a continuous function on U_δ. Hence g is a continuous function on the unit sphere in $\mathcal{R}^s$. Also

$$\frac{1}{\delta}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{g}\left(\frac{\theta - \theta_0}{\delta}\right) = -\frac{1}{\delta}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{v}(x, \theta)$$

$$\leqq -\frac{1}{\delta}\mu_0 \| \theta - \theta_0 \|^2 + \delta\kappa_3 \| \theta - \theta_0 \|,$$

if $\theta \varepsilon U_\delta$, since $\mathbf{B}_{\theta_0}$ is positive definite with minimum latent root μ_0 and, by Lemma 1, $\| v(x, \theta) \| < \kappa_3$ when $\theta \varepsilon U_\delta$. Hence for every θ such that $\| \theta - \theta_0 \| = \delta$, we have

$$\frac{1}{\delta}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{g}\left(\frac{\theta - \theta_0}{\delta}\right) \leqq \delta(\delta\kappa_3 - \mu_0)$$

$$< 0, \quad \text{if} \quad \delta < \frac{\mu_0}{\kappa_3}.$$

Hence if $\delta < \mu_0/\kappa_3$, it follows by Lemma 2 that there exists a point $\hat{\theta}(x)$ such that $\hat{\theta}(x) \varepsilon U_\delta$ and $g((\hat{\theta}(x) - \theta_0) / \delta) = 0$, i.e., $\hat{\theta}(x)$ is a solution of (3.6). The result follows by application of Lemma 1.

4. Existence of a maximum of $L(x, \theta)$. In this paragraph we will show that for sufficiently small δ, if $x \varepsilon X_n$, any solution of (3.6) in U_δ maximises $L(x, \theta)$ subject to the condition $h(\theta) = 0$.

We suppose that $x \varepsilon X_n$, that δ is small enough for Theorem 1 to apply and that $\hat{\theta}(x)$, written $\hat{\theta}$ for typographical brevity, is a solution in U_δ of (3.6). We let θ be a point in a neighbourhood of $\hat{\theta}$ contained in U_δ, such that $h(\theta) = 0$. (Such a neighbourhood exists since $\hat{\theta}$ is an interior point of U_δ.) Then by expanding $L(x, \theta)$ about $\hat{\theta}$ we have

$$(4.1) \qquad L(x, \theta) - L(x, \hat{\theta}) = \mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{M}_{x,\theta^*}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

where $\mathbf{M}_{x,\theta^*} = (\partial^2 L(x, \theta^*) / \partial\theta_i\partial\theta_j)$ and $\theta^* \varepsilon U_\delta$.

We now consider separately the two terms in the right hand side of (4.1). By (3.12)

$$\mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = \mathbf{l}'(x, \hat{\theta})\mathbf{Q}_{\hat{\theta}}'\mathbf{H}_{\hat{\theta}}'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Now

$$\mathbf{0} = \mathbf{h}(\theta) - \mathbf{h}(\hat{\theta}) = \mathbf{H}_{\hat{\theta}}'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \mathbf{r}(\theta),$$

where, because of $\mathcal{H}$2, by the same argument as was applied to $v^{(2)}(\theta)$ in (3.2),

(4.2) $$\| r(\theta)\| < s^3\kappa_2 \| \theta - \hat{\theta} \|^2.$$

Hence

(4.3) $$\mathbf{l}'(x, \hat{\theta})(\theta - \hat{\theta}) = -[Q_{\hat{\theta}}\mathbf{l}(x, \hat{\theta})]'\mathbf{r}(\theta).$$

By (3 8)

$$\frac{1}{n}\mathbf{l}(x, \hat{\theta}) = -\mathbf{B}_{\theta_0}(\hat{\theta} - \theta_0) + \mathbf{v}^{(6)}(x, \hat{\theta}),$$

and so

$$\frac{1}{n}\| \boldsymbol{l}(x, \hat{\theta}) \| < \kappa_4\delta + \kappa_5\delta^2, \quad \text{since} \quad \hat{\theta} \, \varepsilon \, U_\delta,$$

where κ_4 is a positive number depending only on the elements of $\mathbf{B}_{\theta_0}$, and, as above, $\kappa_5 = 1 + s^2 + s^3\kappa_1$. Also the elements of $Q_{\hat{\theta}}$ are bounded by a number independent of δ, since $\hat{\theta} \, \varepsilon \, U_\alpha$. Hence

(4.4) $$\frac{1}{n}\| Q_{\hat{\theta}}\, l(x, \hat{\theta}) \| < \kappa_6\delta + \kappa_7\delta^2,$$

where κ_6, κ_7 are positive numbers independent of δ. From (4.2), (4.3) and (4.4) it follows that

(4.5) $$\frac{1}{n}| \mathbf{l}'(x, \hat{\theta})(\theta - \hat{\theta}) | < (\kappa_6\delta + \kappa_7\delta^2)s^3\kappa_2 \| \theta - \hat{\theta} \|^2.$$

We now consider the second term of (4.1). By expanding the elements of $\mathbf{M}_{x,\theta^*}$ about θ_0 we find that

$$\frac{1}{n}\mathbf{M}_{x,\theta^*} = \frac{1}{n}\mathbf{M}_{x,\theta_0} + \mathbf{m}^*_{x,\theta^*},$$

where, as is easily shown using $\mathfrak{X}4$, the moduli of the elements of the matrix $\mathbf{m}^*_{x,\theta^*}$ are less than $2s\kappa_1\delta$. Also by $\mathfrak{X}3$,

$$\frac{1}{n}\mathbf{M}_{x,\theta_0} = -\mathbf{B}_{\theta_0} + \delta\mathbf{m}_{x,\theta_0},$$

and so

$$\frac{1}{n}\mathbf{M}_{x,\theta^*} = -\mathbf{B}_{\theta_0} + \delta\mathbf{m},$$

say, where $\mathbf{m}$ is a matrix whose elements are bounded by a number independent of δ. Hence

(4.6) $$\frac{1}{2n}(\theta - \hat{\theta})'\mathbf{M}_{x,\theta^*}(\theta - \hat{\theta}) = -\frac{1}{2}(\theta - \hat{\theta})'\mathbf{B}_{\theta_0}(\theta - \hat{\theta})$$

$$+ \frac{1}{2}\delta(\theta - \hat{\theta})'\mathbf{m}(\theta - \hat{\theta}) < -\frac{1}{2}\mu_0 \| \theta - \hat{\theta} \|^2 + \kappa_8\delta \| \theta - \hat{\theta} \|^2,$$

since B_{θ_0} is positive definite with minimum latent root μ_0, and the elements of m are bounded. Here κ_8 is a positive number depending only on the elements of m. Using (4.5) and (4.6) in (4.1) we find that there exist positive numbers κ_9, κ_{10}, independent of δ, such that

$$\frac{1}{n}[L(x, \theta) - L(x, \hat{\theta})] < \left(-\frac{1}{2}\mu_0 + \kappa_9\delta + \kappa_{10}\delta^2\right) \| \theta - \hat{\theta} \|^2.$$

It follows that if δ is sufficiently small then $L(x, \theta) < L(x, \hat{\theta})$, i.e., $L(x, \hat{\theta})$ is a maximum value of $L(x, \theta)$ subject to $h(\theta) = 0$.

We have thus established the fact that, if the conditions $\mathfrak{F}$ and $\mathfrak{H}$ are satisfied, there exists a consistent maximum likelihood estimator $\hat{\theta}$ of θ_0 satisfying the condition $h(\hat{\theta}) = 0$.

5. Asymptotic distributions. We return now to consideration of (3.1) and (3.2). We suppose that $x \in X_n$ and that $\hat{\theta}(x)$, $\hat{\lambda}(x)$ is a solution of these equations with $\hat{\theta}(x) \in U_\delta$, δ being small enough for such a solution to exist. Then, considering the equations from a slightly different viewpoint we have,

$$(5.1) \qquad \frac{1}{n}\mathrm{l}(x, \theta_0) - [\mathrm{B}_{\theta_0} + \hat{\mathrm{b}}(x)][\hat{\theta}(x) - \theta_0] + [\mathrm{H}_{\theta_0} + \hat{\mathrm{h}}(x)]\frac{1}{n}\hat{\lambda}(x) = 0,$$

$$(5.2) \qquad [\mathrm{H}'_{\theta_0} + \hat{\mathrm{h}}^*(x)][\hat{\theta}(x) - \theta_0] = 0,$$

where $\hat{\mathrm{b}}(x)$, $\hat{\mathrm{h}}(x)$ and $\hat{\mathrm{h}}^*(x)$ are matrices whose elements tend to 0 as δ (and hence $\| \hat{\theta}(x) - \theta_0 \|) \to 0$. We now prove the following lemma.

Lemma 3. *The partitioned matrix*

$$\begin{bmatrix} \mathrm{B}_{\theta_0} & -\mathrm{H}_{\theta_0} \\ -\mathrm{H}'_{\theta_0} & 0 \end{bmatrix}$$

is non-singular.

Proof. For brevity we omit the suffix θ_0. Then we wish to find a matrix

$$\begin{bmatrix} \mathrm{P} & \mathrm{Q} \\ \mathrm{Q}' & \mathrm{R} \end{bmatrix}$$

such that, in the usual notation,

$$\begin{bmatrix} \mathrm{B} & -\mathrm{H} \\ -\mathrm{H}' & 0 \end{bmatrix}\begin{bmatrix} \mathrm{P} & \mathrm{Q} \\ \mathrm{Q}' & \mathrm{R} \end{bmatrix} = \begin{bmatrix} \mathrm{I}_s & 0 \\ 0 & \mathrm{I}_r \end{bmatrix}$$

and this requires

$$(5.3) \qquad \mathrm{BP} - \mathrm{HQ}' = \mathrm{I}_s,$$

$$(5.4) \qquad \mathrm{BQ} - \mathrm{HR} = 0,$$

$$(5.5) \qquad \mathrm{H}'\mathrm{P} = 0,$$

$$(5.6) \qquad -\mathrm{H}'\mathrm{Q} = \mathrm{I}_r.$$

These equations are easily solved since B is positive definite and H is of rank

r so that $H'B^{-1}H$ is non-singular. We obtain

$$R = -(H'B^{-1}H)^{-1},$$
$$Q = -BH(H'B^{-1}H)^{-1},$$
$$P = B^{-1}[I_s - H(H'B^{-1}H)^{-1}H'B^{-1}].$$

We note at this stage, though we do not require this result immediately, that the matrix P has rank $s - r$. For, from (5.5) since rank $(H') = r$, rank $(P) \leqq s - r$. While from (5.3) we have $s =$ rank $(P - HQ') \leqq$ rank $(P) +$ rank $(HQ') \leqq$ rank $(P) + r$, and so rank $(P) \geqq s - r$.

We return now to equations (5.1) and (5.2). If δ is sufficiently small then the matrix

$$\begin{bmatrix} B_{\theta_0} + \hat{b}(x) & -[H_{\theta_0} + \hat{h}(x)] \\ -[H'_{\theta_0} + \hat{h}^*(x)] & 0 \end{bmatrix}$$

also will be non-singular and we will write

$$\begin{bmatrix} B_{\theta_0} + \hat{b}(x) & -[H_{\theta_0} + \hat{h}(x)] \\ -[H'_{\theta_0} + \hat{h}^*(x)] & 0 \end{bmatrix}^{-1} = \begin{bmatrix} \hat{P}(x) & \hat{Q}_1(x) \\ \hat{Q}_2(x) & \hat{R}(x) \end{bmatrix}.$$

Hence, from (5.1) and (5.2), for sufficiently small δ, if $x \,\varepsilon\, X_n$, we have

$$\begin{bmatrix} \hat{\theta}(x) - \theta_0 \\ \frac{1}{n}\hat{\lambda}(x) \end{bmatrix} = \begin{bmatrix} \hat{P}(x) & \hat{Q}_1(x) \\ \hat{Q}_2(x) & \hat{R}(x) \end{bmatrix} \begin{bmatrix} \frac{1}{n} l(x, \theta_0) \\ 0 \end{bmatrix}. \tag{5.7}$$

If the functions $\hat{\theta}$ and $\hat{\lambda}$ were defined for the whole of $\mathfrak{R}^n$ we could now discuss immediately the asymptotic distribution of these functions. However this is not necessarily so, and we go through the formality of extending the definition of these functions to the whole of $\mathfrak{R}^n$. We will then show that the random variables thus defined are asymptotically normally distributed and, in this sense, we may say that a consistent maximum likelihood estimator [illegible] of θ_0 is asymptotically normally distributed.

We let (δ_m), (ϵ_m) be decreasing sequences of positive numbers [illegible] $\epsilon_1 < 1$, $\delta_1 < \mu_0/\kappa_3$ (see Theorem 1), and $\delta_m \to 0$ and $\epsilon_m \to 0$ [illegible] ∞. [illegible] then define an increasing sequence (n_m) of integers such that [illegible] exists a set in $\mathfrak{R}^n$ with the properties $\mathfrak{X}1$ to $\mathfrak{X}4$ for $\epsilon = \epsilon_m$ [illegible] $m = 1, 2, \cdots$, if $n_m \leqq n < n_{m+1}$ we choose a set X_n [illegible] to $\mathfrak{X}4$ for $\epsilon = \epsilon_m$ and $\delta = \delta_m$. Hence $\Pr\{X_n\} \to 1$ as $n \to$ [illegible] $< n_{m+1}$ and $x \,\varepsilon\, X_n$, the likelihood equations (2.1) and [illegible] $\hat{\theta}_n(x)$, $\hat{\lambda}_n(x)$ such that $\| \hat{\theta}_n(x) - \theta_0 \| < \delta_m$. Moreover [illegible] $\hat{\theta}_n(x)$ is a *maximum* likelihood estimate of θ_0, by [illegible] nition of $\hat{\theta}_n$ and $\hat{\lambda}_n$ to $\mathfrak{R}^n$ by letting

$$\begin{bmatrix} \hat{\theta}_n(x) - \theta_0 \\ \frac{1}{n}\hat{\lambda}_n(x) \end{bmatrix} = \begin{bmatrix} P & Q \\ Q' & R \end{bmatrix} \begin{bmatrix} \frac{1}{n} l(x, \text{[illegible]} \\ 0 \end{bmatrix} \text{[illegible]}$$

We have thus defined sequences $(\hat{\theta}_n)$, $(\hat{\lambda}_n)$, $n = n_m, n_{m+1}, \cdots$ of random variables which have the property that θ_n converges in probability to θ_0 and with probability tending to 1 as $n \to \infty$, $\hat{\theta}_n$, $\hat{\lambda}_n$ satisfy the likelihood equations (2.1) and (2.2).

THEOREM 2. *The random variables* $n^{1/2}(\hat{\theta}_n - \theta_0)$, $n^{-1/2}\hat{\lambda}_n$ *are asymptotically jointly normally distributed with variance-covariance matrix*

$$\begin{bmatrix} P & 0 \\ 0 & -R \end{bmatrix}.$$

PROOF. If $x \varepsilon X_n$, we define $\hat{P}(x) = P$, $\hat{Q}_1(x) = Q$, $\hat{Q}_2(x) = Q'$ and $\hat{R}(x) = R$. Then for sufficiently large n, by (5.7) we may write

$$\begin{bmatrix} \sqrt{n}(\hat{\theta}_n - \theta_0) \\ \frac{1}{\sqrt{n}}\hat{\lambda}_n \end{bmatrix} = \begin{bmatrix} \hat{P} & \hat{Q}_1 \\ \hat{Q}_2 & \hat{R} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{n}} l(\cdot, \theta_0) \\ 0 \end{bmatrix}.$$

The elements of the matrix

$$\begin{bmatrix} \hat{P} & \hat{Q}_1 \\ \hat{Q}_2 & \hat{R} \end{bmatrix}$$

are random variables which converge in probability to the corresponding elements of the matrix

$$\begin{bmatrix} P & Q \\ Q' & R \end{bmatrix},$$

since in (5.1) and (5.2) $\hat{\mathfrak{b}}$, $\hat{\mathfrak{h}}$ and $\hat{\mathfrak{h}}^*$ tend to 0 as $\delta \to 0$. Also the s-dimensional random variable $n^{-1/2}\ell(\cdot, \theta_0)$ is asymptotically normally distributed with zero mean and variance-covariance matrix B_{θ_0} (Cramér [1]), and the $(s + r)$-dimensional random variable $(n^{-1/2}\ell(\cdot, \theta_0), 0)$ is asymptotically normally distributed with zero mean and variance-covariance matrix

$$\begin{bmatrix} B_{\theta_0} & 0 \\ 0 & 0 \end{bmatrix}.$$

It follows by an extension, to a multi-dimensional random variable, of a theorem of Cramér [2], that $\sqrt{n}(\hat{\theta}_n - \theta_0)$, $n^{-1/2}\hat{\lambda}_n$ are jointly asymptotically normally distributed with zero mean and variance-covariance matrix.

$$\begin{bmatrix} P & Q \\ Q' & R \end{bmatrix} \begin{bmatrix} B_{\theta_0} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P & Q \\ Q' & R \end{bmatrix} = \begin{bmatrix} PB_{\theta_0}P & PB_{\theta_0}Q \\ Q'B_{\theta_0}P & Q'B_{\theta_0}Q \end{bmatrix}.$$

(We omit details of the proof of this extension though this result, in contrast to Cramér's result for real-valued random variables, is best obtained by considering characteristic functions). Now from (5.3), $PB_{\theta_0}P - PH_{\theta_0}Q' = P$. Since P is symmetric, $PH_{\theta_0} = P'H_{\theta_0} = 0$ by (5.5). Hence $PB_{\theta_0}P = P$. Similarly $PB_{\theta_0}Q = 0$ and $Q'B_{\theta_0}Q = -R$.

This completes the proof of the Theorem. We note, however, that, as might be expected, the asymptotic normal distribution of the s-dimensional random variable $n^{1/2}(\hat{\theta}_n - \theta_0)$ is improper, being by the note in Lemma 3 of rank $s - r$.

6. Numerical solution of likelihood equations. In this section we will discuss an iterative procedure for solving (2.1) and (2.2) numerically, which yields an estimate of the matrices $\mathbf{P}$ and $\mathbf{R}$.

In any practical situation we do not know θ_0, and the only way in which we can verify that the conditions $\mathfrak{F}$ and $\mathfrak{H}$ are satisfied is to find that, for every θ belonging *to some set U, in which we know θ_0 lies*, the *following conditions* $\mathfrak{F}'$, $\mathfrak{H}'$ are satisfied.

$\mathfrak{F}'1$, $\mathfrak{F}'2$. For every $\theta \varepsilon U$, $\mathfrak{F}1$ and $\mathfrak{F}2$ are satisfied.

$\mathfrak{F}'3$ For every $\theta \varepsilon U$ and $i, j, k = 1, 2, \cdots, s$,

$$\left| \frac{\partial^3 \log f(t, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < F_3(t)$$

and

$$\int_{-\infty}^{\infty} F_3(t) f(t, \theta)\, dt \leqq \kappa_1',$$

a finite number.

$\mathfrak{F}'4$. For every $\theta \varepsilon U$,

$$b_{ij} = \int_{-\infty}^{\infty} \frac{\partial \log f(t, \theta)}{\partial\theta_i} \frac{\partial \log f(t, \theta)}{\partial\theta_j} f(t, \theta)\, dt,$$

$i, j = 1, 2, \cdots, s$, are finite, the matrix $\mathbf{B}_\theta = (b_{ij}(\theta))$ is positive definite and, if μ_θ is the minimum latent root of $\mathbf{B}_\theta$, then $\mu_\theta \geqq \mu_0'$ where μ_0' is a given number *greater than* 0.

$\mathfrak{H}'1$, $\mathfrak{H}'2$ For every $\theta \varepsilon U$, $\mathfrak{H}1$ and $\mathfrak{H}2$ are satisfied

$\mathfrak{H}'3$ For every $\theta \varepsilon U$, $\mathbf{H}_\theta$ is of rank r.

The conditions $\mathfrak{F}'$ are a straightforward generalization of Cramér's conditions [2].

We will now assume that the conditions $\mathfrak{F}'$ and $\mathfrak{H}'$ are satisfied, that x is such that the likelihood equations (2.1) and (2.2) have a solution $\hat{\theta}(x), \hat{\lambda}(x)$ and that $\theta^{(1)}$ is an initial approximation to $\hat{\theta}(x)$ such that $\| \theta^{(1)} - \hat{\theta}(x) \|$ is small. Then to a first order of approximation

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{l}(x, \theta^{(1)}) + \mathbf{M}_{x,\theta^{(1)}}(\hat{\theta} - \theta^1),$$
$$\mathbf{h}(\hat{\theta}) = \mathbf{h}(\theta^{(1)}) + \mathbf{H}_{\theta^{(1)}}(\hat{\theta} - \theta^{(1)}).$$

Also if n is large, $(1/n)\hat{\lambda}(x)$ is near $\mathbf{0}$ for "most" x. We assume that x is a point for which $(1/n)\lambda(x)$ is near $\mathbf{0}$. Then we also have to a first order of approximation

We have thus defined sequences $(\hat{\theta}_n)$, $(\hat{\lambda}_n)$, $n = n_m, n_{m+1}, \cdots$ of random variables which have the property that θ_n converges in probability to θ_0 and with probability tending to 1 as $n \to \infty$, $\hat{\theta}_n$, $\hat{\lambda}_n$ satisfy the likelihood equations (2.1) and (2.2).

THEOREM 2. *The random variables* $n^{1/2}(\hat{\theta}_n - \theta_0)$, $n^{-1/2}\hat{\lambda}_n$ *are asymptotically jointly normally distributed with variance-covariance matrix*

$$\begin{bmatrix} \mathrm{P} & 0 \\ 0 & -\mathrm{R} \end{bmatrix}.$$

PROOF. If $x \varepsilon X_n$, we define $\hat{\mathrm{P}}(x) = \mathrm{P}$, $\hat{\mathrm{Q}}_1(x) = \mathrm{Q}$, $\hat{\mathrm{Q}}_2(x) = \mathrm{Q}'$ and $\hat{\mathrm{R}}(x) = \mathrm{R}$. Then for sufficiently large n, by (5.7) we may write

$$\begin{bmatrix} \sqrt{n}(\hat{\theta}_n - \theta_0) \\ \frac{1}{\sqrt{n}} \hat{\lambda}_n \end{bmatrix} = \begin{bmatrix} \hat{\mathrm{P}} & \hat{\mathrm{Q}}_1 \\ \hat{\mathrm{Q}}_2 & \hat{\mathrm{R}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{n}} \mathrm{l}(\cdot, \theta_0) \\ 0 \end{bmatrix}.$$

The elements of the matrix

$$\begin{bmatrix} \hat{\mathrm{P}} & \hat{\mathrm{Q}}_1 \\ \hat{\mathrm{Q}}_2 & \hat{\mathrm{R}} \end{bmatrix}$$

are random variables which converge in probability to the corresponding elements of the matrix

$$\begin{bmatrix} \mathrm{P} & \mathrm{Q} \\ \mathrm{Q}' & \mathrm{R} \end{bmatrix},$$

since in (5.1) and (5.2) $\hat{\mathrm{b}}$, $\hat{\mathrm{h}}$ and $\hat{\mathrm{h}}^*$ tend to $\mathbf{0}$ as $\delta \to 0$. Also the s-dimensional random variable $n^{-1/2}\ell(\cdot, \theta_0)$ is asymptotically normally distributed with zero mean and variance-covariance matrix B_{θ_0} (Cramér [1]), and the $(s + r)$-dimensional random variable $(n^{-1/2}\ell(\cdot, \theta_0), 0)$ is asymptotically normally distributed with zero mean and variance-covariance matrix

$$\begin{bmatrix} \mathrm{B}_{\theta_0} & 0 \\ 0 & 0 \end{bmatrix}.$$

It follows by an extension, to a multi-dimensional random variable, of a theorem of Cramér [2], that $\sqrt{n}(\hat{\theta}_n - \theta_0)$, $n^{-1/2}\hat{\lambda}_n$ are jointly asymptotically normally distributed with zero mean and variance-covariance matrix.

$$\begin{bmatrix} \mathrm{P} & \mathrm{Q} \\ \mathrm{Q}' & \mathrm{R} \end{bmatrix} \begin{bmatrix} \mathrm{B}_{\theta_0} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathrm{P} & \mathrm{Q} \\ \mathrm{Q}' & \mathrm{R} \end{bmatrix} = \begin{bmatrix} \mathrm{PB}_{\theta_0}\mathrm{P} & \mathrm{PB}_{\theta_0}\mathrm{Q} \\ \mathrm{Q}'\mathrm{B}_{\theta_0}\mathrm{P} & \mathrm{Q}'\mathrm{B}_{\theta_0}\mathrm{Q} \end{bmatrix}.$$

(We omit details of the proof of this extension though this result, in contrast to Cramér's result for real-valued random variables, is best obtained by considering characteristic functions). Now from (5.3), $\mathrm{PB}_{\theta_0}\mathrm{P} - \mathrm{PH}_{\theta_0}\mathrm{Q}' = \mathrm{P}$. Since P is symmetric, $\mathrm{PH}_{\theta_0} = \mathrm{P}'\mathrm{H}_{\theta_0} = \mathbf{0}$ by (5.5). Hence $\mathrm{PB}_{\theta_0}\mathrm{P} = \mathrm{P}$. Similarly $\mathrm{PB}_{\theta_0}\mathrm{Q} = \mathbf{0}$ and $\mathrm{Q}'\mathrm{B}_{\theta_0}\mathrm{Q} = -\mathrm{R}$.

This completes the proof of the Theorem. We note, however, that, as might be expected, the asymptotic normal distribution of the s-dimensional random variable $n^{1/2}(\hat{\theta}_n - \theta_0)$ is improper, being by the note in Lemma 3 of rank $s - r$.

6. Numerical solution of likelihood equations. In this section we will discuss an iterative procedure for solving (2.1) and (2.2) numerically, which yields an estimate of the matrices **P** and **R**.

In any practical situation we do not know θ_0, and the only way in which we can verify that the conditions $\mathfrak{F}$ and $\mathfrak{H}$ are satisfied is to find that, for every θ belonging to some set U, *in which we know* θ_0 *lies*, the following conditions $\mathfrak{F}'$, $\mathfrak{H}'$ are satisfied.

$\mathfrak{F}'1$, $\mathfrak{F}'2$. For every $\theta \varepsilon U$, $\mathfrak{F}1$ and $\mathfrak{F}2$ are satisfied.

$\mathfrak{F}'3$ For every $\theta \varepsilon U$ and $i, j, k = 1, 2, \cdots, s$,

$$\left| \frac{\partial^3 \log f(t, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < F_3(t)$$

and

$$\int_{-\infty}^{\infty} F_3(t) f(t, \theta)\, dt \leqq \kappa_1',$$

a finite number.

$\mathfrak{F}'4$. For every $\theta \varepsilon U$,

$$b_{ij} = \int_{-\infty}^{\infty} \frac{\partial \log f(t, \theta)}{\partial\theta_i} \frac{\partial \log f(t, \theta)}{\partial\theta_j} f(t, \theta)\, dt,$$

$i, j = 1, 2, \cdots, s$, are finite, the matrix $\mathbf{B}_\theta = (b_{ij}(\theta))$ is positive definite and, if μ_θ is the minimum latent root of $\mathbf{B}_\theta$, then $\mu_\theta \geqq \mu_0'$ where μ_0' is a given number greater than 0.

$\mathfrak{H}'1$, $\mathfrak{H}'2$. For every $\theta \varepsilon U$, $\mathfrak{H}1$ and $\mathfrak{H}2$ are satisfied.

$\mathfrak{H}'3$ For every $\theta \varepsilon U$, $\mathbf{H}_\theta$ is of rank r

The conditions $\mathfrak{F}'$ are a straightforward generalization of Cramér's conditions [2].

We will now assume that the conditions $\mathfrak{F}'$ and $\mathfrak{H}'$ are satisfied, that x is such that the likelihood equations (2.1) and (2.2) have a solution $\hat{\theta}(x), \hat{\lambda}(x)$ and that $\theta^{(1)}$ is an initial approximation to $\hat{\theta}(x)$ such that $\| \theta^{(1)} - \hat{\theta}(x) \|$ is small. Then to a first order of approximation

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{l}(x, \theta^{(1)}) + \mathbf{M}_{x,\theta^{(1)}}(\hat{\mathbf{\theta}} - \mathbf{\theta}^1),$$

$$\mathbf{h}(\hat{\theta}) = \mathbf{h}(\theta^{(1)}) + \mathbf{H}_{\theta^{(1)}}(\hat{\mathbf{\theta}} - \mathbf{\theta}^{(1)}).$$

Also if n is large, $(1/n)\hat{\lambda}(x)$ is near **0** for "most" x. We assume that x is a point for which $(1/n)\lambda(x)$ is near **0**. Then we also have to a first order of approximation

$$H_{\hat\theta}\frac{1}{n}\hat\lambda = H_{\theta^{(1)}}\frac{1}{n}\hat\lambda.$$

ice $\hat\theta(x)$, $\hat\lambda(x)$ satisfy (2.1) and (2.2) then, approximately, we have

$$\begin{bmatrix} -\frac{1}{n}M_{x,\theta^{(1)}} & -H_{\theta^{(1)}} \\ -H'_{\theta^{(1)}} & 0 \end{bmatrix}\begin{bmatrix} \hat\theta - \theta^{(1)} \\ \frac{1}{n}\hat\lambda \end{bmatrix} = \begin{bmatrix} \frac{1}{n}l(x, \theta^{(1)}) \\ h(\theta^{(1)}) \end{bmatrix}. \quad (6.1)$$

The normal situation, if n is large, is that $\hat\theta(x)$ is near θ_0. Consequently since is near $\hat\theta(x)$ the matrix $-(1/n)M_{x,\theta^{(1)}}$ approximates $-(1/n)M_{x,\theta_0}$ which in n approximates B_{θ_0}. Then $B_{\theta^{(1)}}$ approximates B_{θ_0} and we propose to replace 1/n)$M_{x,\theta^{(1)}}$ in (6.1) by $B_{\theta^{(1)}}$, and to obtain a correction to $\theta^{(1)}$, and an initial proximation to $(1/n)\hat\lambda$, by solving the equation

$$\begin{bmatrix} B_{\theta^{(1)}} & -H_{\theta^{(1)}} \\ -H'_{\theta^{(1)}} & 0 \end{bmatrix}\begin{bmatrix} \hat\theta - \theta^{(1)} \\ \frac{1}{n}\hat\lambda \end{bmatrix} = \begin{bmatrix} \frac{1}{n}l(x, \theta^{(1)}) \\ h(\theta^{(1)}) \end{bmatrix}. \quad (6.2)$$

The idea of replacing $-(1/n)M_{x,\theta^{(1)}}$ by $B_{\theta^{(1)}}$ is not original though the authors not know where it originated.

Because of $\mathcal{F}'4$, $\mathcal{H}'3$, by Lemma 3, the matrix

$$\begin{bmatrix} B_{\theta^{(1)}} & -H_{\theta^{(1)}} \\ -H'_{\theta^{(1)}} & 0 \end{bmatrix}$$

non-singular and we will denote its inverse by

$$\begin{bmatrix} P_1 & Q_1 \\ Q'_1 & R_1 \end{bmatrix}.$$

define $\theta^{(2)}$, $\lambda^{(2)}$ by

$$\begin{bmatrix} \theta^{(2)} \\ \frac{1}{n}\lambda^{(2)} \end{bmatrix} = \begin{bmatrix} \theta^{(1)} \\ 0 \end{bmatrix} + \begin{bmatrix} P_1 & Q_1 \\ Q'_1 & R_1 \end{bmatrix}\begin{bmatrix} \frac{1}{n}l(x, \theta^{(1)}) \\ h(\theta^{(1)}) \end{bmatrix}$$

l, more generally, $\theta^{(r)}$, $\lambda^{(r)}$ by (with the obvious definition of P_{r-1}, Q_{r-1} l R_{r-1}),

$$\begin{bmatrix} \theta^{(r)} \\ \frac{1}{n}\lambda^{(r)} \end{bmatrix} = \begin{bmatrix} \theta^{(r-1)} \\ 0 \end{bmatrix} + \begin{bmatrix} P_{r-1} & Q_{r-1} \\ Q'_{r-1} & R_{r-1} \end{bmatrix}\begin{bmatrix} \frac{1}{n}l(x, \theta^{(r-1)}) \\ h(\theta^{(r-1)}) \end{bmatrix}.$$

If the sequences $(\theta^{(r)})$, $(\lambda^{(r)})$ converge then they converge to a solution of the elihood equations, as is easily verified. We do not attempt to give rigorous nditions under which these sequences do converge. However the fact that may expect them to converge in most practical situations follows from the uristic argument leading to (6.2).

We have thus established an iterative procedure for solving the likelihood equations. The heaviest part of the computation involved in this method is the inversion of a matrix and computation will normally be reduced by considering the sequences $(\hat{\theta}^{(r)})$, $(\hat{\lambda}^{(r)})$ defined by

$$\begin{bmatrix} \hat{\theta}^{(r)} \\ \frac{1}{n}\hat{\lambda}^{(r)} \end{bmatrix} = \begin{bmatrix} \hat{\theta}^{(r-1)} \\ \frac{1}{n}\hat{\lambda}^{(r-1)} \end{bmatrix} + \begin{bmatrix} P_1 & Q_1 \\ Q_1' & R_1 \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \hat{\theta}^{(r-1)}) + \mathbf{H}_{\hat{\theta}^{(r-1)}} \frac{1}{n}\hat{\lambda}^{(r-1)} \\ \mathbf{h}(\hat{\theta}^{(r-1)}) \end{bmatrix}$$

$r = 1, 2, \cdots$, where $\hat{\theta}^{(0)} = \theta^{(2)}$ and $\hat{\lambda}^{(0)} = \lambda^{(2)}$. Again if these sequences converge, they converge to a solution of the likelihood equations since

$$\begin{bmatrix} P_1 & Q_1 \\ Q_1' & R_1 \end{bmatrix}$$

is non-singular. And again we do not attempt to give conditions under which they do converge. The main justifications we put forward for this computational procedure are

(i) the similarity between this method and Newton's method, and

(ii) the fact that similar modifications of Newton's method have been used successfully elsewhere, for example in probit analysis [3]. The main advantage of this method of solving the likelihood equations is that it involves inversion of only one matrix.

7. Tests of the model. In a situation such as is outlined in §1 two natural questions arise in practice regarding the adequacy of the model introduced to describe an experimental situation.

(i) Does the true parameter point θ_0 satisfy the condition $h(\theta_0) = 0$?

(ii) Is the true parameter point some hypothetical point θ^* such that

$$h(\theta^*) = 0?$$

And this is the natural order for these questions since the second would be asked only if the first were answered in the affirmative. We now propose a procedure for answering these questions in this order.

(i) The most natural approach to the first question would be as follows. We would calculate an unrestrained maximum likelihood estimate $\hat{\theta}_u(x)$ of θ_0, and for $\hat{\theta}_u(x)$ we would have $l(x, \hat{\theta}_u(x)) \doteq 0$. If $h(\hat{\theta}_u(x))$ were in some sense "near enough" $0 \in \mathfrak{R}^r$ then we would decide that in fact $h(\theta_0) = 0$. Dually, we might calculate a maximum likelihood estimate $\hat{\theta}(x)$ subject to the restraint

$$h(\hat{\theta}(x)) = 0$$

and then decide that $h(\theta_0) = 0$ if $l(x, \hat{\theta}(x))$ were "near enough" $0 \in \mathfrak{R}^s$. And the test we propose is based on the second possibility. We note that, by (2.1),

$$\mathbf{H}_{\hat{\theta}}\hat{\lambda}(x) = -\mathbf{l}(x, \hat{\theta}(x))$$

and it seems reasonable to decide that $h(\theta_0) = 0$ if $\hat{\lambda}(x)$ is in some sense 'near enough' $0 \in \mathfrak{R}^r$.

We have seen in Theorem 2 that when $h(\theta_0) = 0$, $n^{-1/2}\hat{\lambda}$ is normally distributed asymptotically with variance-covariance matrix $-\mathrm{R}$, which is of rank r. Consequently $-(1/n)\hat{\lambda}'\mathrm{R}^{-1}\hat{\lambda}$ is asymptotically distributed as χ^2 with r degrees of freedom, when $h(\theta_0) = 0$, and, in obvious notation, $-(1/n)\hat{\lambda}'\mathrm{R}_{\hat{\theta}}^{-1}\hat{\lambda}$ also is approximately, for large n, distributed as χ^2 with r degrees of freedom. We propose to choose as a region of acceptance of the hypothesis that $h(\theta_0) = 0$ the set of x for which

$$-\frac{1}{n}\hat{\lambda}'(x)\mathrm{R}_{\hat{\theta}(x)}^{-1}\hat{\lambda}(x) \leqq k,$$

where k is determined by

$$\Pr\{\chi_{[r]}^2 \leqq k\} = 0.95.$$

This gives a test of size 95% of the hypothesis that $h(\theta_0) = 0$.

(ii) The natural corollary of using the asymptotic distribution of $\hat{\lambda}$ in this way is to use the asymptotic distribution of $\hat{\theta}$ as established in Theorem 2 to answer the second question. If $\theta^* = \theta_0$ then $n(\hat{\theta} - \theta^*)'\mathrm{B}_{\theta^*}(\hat{\theta} - \theta^*)$ is approximately distributed as χ^2 with $s - r$ degrees of freedom if n is large. This is easily established by noting that a consequence of equations (5.3)–(5.6) is that $\mathrm{B}^{-1} = \mathrm{PBP} - \mathrm{QR}^{-1}\mathrm{Q}'$, and hence that

$$\frac{1}{n}\mathrm{l}'\mathrm{B}^{-1}\mathrm{l} = n(\hat{\theta} - \theta_0)'\mathrm{B}(\hat{\theta} - \theta_0) - \frac{1}{n}\hat{\lambda}'\mathrm{R}^{-1}\hat{\lambda}.$$

We use this fact as in the previous paragraph to establish a region of acceptance of the hypothesis that the true parameter point is θ^*.

Here no attempt is made to justify this test on other than an intuitive basis. Since the Lagrangian multiplier test seems to be of wide applicability and of considerable importance in practical statistics, it will be fully discussed both from the theoretical and practical points of view in subsequent papers.

REFERENCES

[1] H. Cramér, "Random variables and probability distributions," *Cambridge University Press*, 1937.
[2] H. Cramér, "Mathematical methods of statistics," *Princeton University Press*, 1949.
[3] D. J. Finney, "Probit analysis," *Cambridge University Press*, 1947.
[4] S. Lefschetz, "Introduction to topology," *Princeton University Press*, 1949.

CONFIDENCE BOUNDS ON VECTOR ANALOGUES OF THE "RATIO OF MEANS" AND THE "RATIO OF VARIANCES" FOR TWO CORRELATED NORMAL VARIATES AND SOME ASSOCIATED TESTS

S. N. Roy and R. F Potthoff

University of North Carolina

1. Summary and Introduction. In this paper confidence bounds are obtained (i) on the ratio of variances of a (possibly) correlated bivariate normal population, and then, by generalization, (ii) on a set of parametric functions of a (possibly) correlated $p + p$ variate normal population, which plays the same role for a $2p$-variate population as the ratio of variances does for the bivariate case, (iii) on the ratio of means of the population indicated in (i), and, by generalization, (iv) on a set of parametric functions of the population indicated in (ii), which plays the same role for this problem as the ratio of means does for the bivariate case. For (i) and (iii) the confidence coefficient is any preassigned $1 - \alpha$ and the distribution involved is the *central* t-distribution, while for (ii) and (iv), the confidence statement in each case is a simultaneous one with a joint confidence coefficient greater than or equal to a preassigned $1 - \alpha$. For (ii) the distribution involved is that of the *central* largest canonical correlation coefficient (squared), and for (iv) the distribution involved is that of the *central* Hotelling's T^2 As far as the authors are aware the results on (ii) and (iv) are new and so perhaps that on (i). But the result on (iii) has been in the field for a long time in various superficially different forms. An important point to keep in mind on these *problems is that, for such confidence bounds and the associated tests of hy*potheses to be physically meaningful, the two variates for the bivariate distribution should be *comparable*. For example, they might refer to the same characteristic of a set of individuals before and after a feed. Likewise, for a $(p + p)$-variate distribution, the p variates of the first set should be comparable to p variates of the second set. For example, they might refer to several characteristics of a set of individuals before and after a treatment. In each case the confidence bounds are obtained by inverting the test of a certain hypothesis, which is indicated at its proper place. Thus, for the $(p + p)$-variate problem, we assume that there are p pairs of comparable variates and it is the pairwise comparison for these p pairs that seems, in this situation, to be physically more meaningful than anything else. Any general bounds that will be obtained in this paper are to be regarded, in a large measure, as a means to this end, although there could conceivably be physical questions, some of which will be illustrated in a later applied paper to be published elsewhere, to which these more general bounds would be pertinent.

Received May 14, 1957; revised February 18, 1958.

2. Confidence bounds for the case (i). Suppose we have a random sample of size $n(>2)$ from a population:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : N\left[\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right].$$

Let us denote the sample means by $\bar{x}_1$, $\bar{x}_2$, and the sample dispersion matrix by

$$\begin{bmatrix} s_1^2 & s_1 s_2 r \\ s_1 s_2 r & s_2^2 \end{bmatrix}.$$

Then for any constant λ, it is easy to check that covariance $(x_1 - \lambda x_2, x_1. + \lambda x_2)$ is $\text{var}(x_1) - \lambda^2 \text{var}(x_2) = \sigma_1^2 - \lambda^2\sigma_2^2$.

This will be zero if $\lambda^2 = \sigma_1^2/\sigma_2^2$. Thus, with a $\lambda^2 = \sigma_1^2/\sigma_2^2$, the variates $x_1 - \lambda x_2$ and $x_1 + \lambda x_2$ will be uncorrelated and hence, denoting by r^* the sample correlation coefficient between these two variates, we have that r^* has the (central) r-distribution, i.e., $\sqrt{n-2}\,r^* / (1 - r^{*2})^{\frac{1}{2}}$ has the (central) t-distribution with d.f. $(n-2)$. But it is easy to check that

$$(2.1)\qquad \begin{aligned} r^* &= \frac{(s_1^2 - \lambda^2 s_2^2)}{[(s_1^2 + \lambda^2 s_2^2 + 2\lambda s_1 s_2 r)(s_1^2 + \lambda^2 s_2^2 - 2\lambda s_1 s_2 r)]^{\frac{1}{2}}} \\ &= \frac{(s_1^2 - \lambda^2 s_2^2)}{[s_1^4 + \lambda^4 s_2^4 + 2\lambda^2 s_1^2 s_2^2(1 - 2r^2)]^{\frac{1}{2}}}. \end{aligned}$$

Now, starting from the statement (with a probability $1 - \alpha$)

$$(2.2)\qquad \sqrt{n-2}\,| r^*/(1 - r^{*2})^{\frac{1}{2}} | \leqq t_{\alpha/2}(n-2), \text{ or } \leqq t_{\alpha/2} \text{ (more simply)},$$

where $t_{\alpha/2}(n-2)$ is the upper $\alpha/2$-point of the (central) t-distribution with d.f. $(n-2)$, and remembering that $\lambda = \sigma_1/\sigma_2$ and substituting from (2.1) for r^* in terms of s_1, s_2 and r, we have, for σ_1^2/σ_2^2, the following confidence equation (2.3) and confidence bounds (2.4) (with a confidence coefficient $1 - \alpha$)

$$(2.3)\qquad \lambda^4 - \left[2 + \frac{4}{n-2} t_{\alpha/2}^2(1 - r^2)\right] \frac{s_1^2}{s_2^2}\lambda^2 + \frac{s_1^4}{s_2^4} \leqq 0,$$

and

$$(2.4)\qquad \begin{aligned} &\frac{s_1^2}{s_2^2}\left[\left(1 + \frac{2}{n-2} t_{\alpha/2}^2(1 - r^2)\right) - \left\{\left(1 + \frac{2}{n-2} t_{\alpha/2}^2(1 - r^2)\right)^2 - 1\right\}^{\frac{1}{2}}\right] \leqq \frac{\sigma_1^2}{\sigma_2^2} \\ &\leqq \frac{s_1^2}{s_2^2}\left[\left(1 + \frac{2}{n-2} t_{\alpha/2}^2(1 - r^2)\right) + \left\{\left(1 + \frac{2}{n-2} t_{\alpha/2}^2(1 - r^2)\right)^2 - 1\right\}^{\frac{1}{2}}\right]. \end{aligned}$$

We notice that $\lambda = \sigma_1/\sigma_2 = 1$ if and only if $\sigma_1 = \sigma_2$.

Notice that (2.2) or (2.3) can be used as an acceptance region for the hypothesis $\sigma_1/\sigma_2 = \lambda$ (any specific value) against the alternative $\sigma_1/\sigma_2 \neq \lambda$. Since the paper was written it has been brought to the notice of the authors that

this region, for the case of $\sigma_1/\sigma_2 = 1$, i.e., for $\sigma_1 = \sigma_2$, has been explicitly given by Walker and Lev [5].

3. Confidence bounds for the case (ii). Suppose we have

$$\mathbf{x}\,(2p \times 1) = \underset{1}{\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}}\begin{matrix} p \\ p \end{matrix} : N\left[\underset{1}{\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}}\begin{matrix} p \\ p \end{matrix},\ \underset{p \qquad p}{\begin{bmatrix} \Sigma_{11} & \Sigma_{11} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix}}\begin{matrix} p \\ p \end{matrix}\right]$$

$$= N[\xi(2p \times 1),\ \ \Sigma(2p \times 2p)] \quad \text{(say)},$$

and a random sample of size $n(> 2p)$ from this population, with a sample dispersion matrix denoted by

$$\underset{p \qquad p}{\begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix}}\begin{matrix} p \\ p \end{matrix} = S(2p \times 2p) \quad \text{(say)}. \tag{3.1}$$

It is well known [3] that we can choose (non-singular) matrices $\mu(p \times p)$ and $\nu(p \times p)$ such that

$$\Sigma_{11} = \mu\mu', \qquad \Sigma_{22} = \nu\nu' \quad \text{and} \quad \Sigma_{12} = \mu D_{\gamma^{1/2}}\nu', \tag{3.2}$$

where γ's, i.e., $\gamma_1, \gamma_2, \cdots, \gamma_p$ are the characteristic roots of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$ and $D_{\gamma^{1/2}}$ is a diagonal matrix whose diagonal elements are $\gamma_1^{\frac{1}{2}}, \cdots, \gamma_p^{\frac{1}{2}}$. It is also well known [3] that these roots are all non-negative, that the number of positive roots is the same as the rank of Σ_{12} and that all the roots are zero if, and only if, $\Sigma_{12} = 0$.

Now introduce a new variate $\mathbf{x}^*$ $(2p \times 1)$ defined by

$$\mathbf{x}^*(2p \times 1) = \begin{matrix} p \\ p \end{matrix}\underset{1}{\begin{bmatrix} \mathbf{x}_1^* \\ \mathbf{x}_2^* \end{bmatrix}} \quad \text{(say)} \quad = A(2p \times 2p)\mathbf{x}(2p \times 1), \tag{3.3}$$

where

$$A(2p \times 2p) = \underset{p \qquad p}{\begin{bmatrix} I & -\mu\nu^{-1} \\ I & \mu\nu^{-1} \end{bmatrix}}\begin{matrix} p \\ p \end{matrix} = \begin{bmatrix} I & -\lambda \\ I & \lambda \end{bmatrix} \quad \text{(say)}. \tag{3.4}$$

Then this $\mathbf{x}^*$ is $N(\xi^*, \Sigma^*)$, where $\xi^* = A\xi$ and

$$\Sigma^* = \begin{bmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\prime} & \Sigma_{22}^* \end{bmatrix} \quad \text{(say)} \quad = A\Sigma A', \tag{3.5}$$

whence we have that

$$\begin{aligned} \Sigma_{11}^* &= 2(\Sigma_{11} - \mu D_{\gamma^{1/2}}\mu'),\ \Sigma_{22}^* = 2(\Sigma_{11} + \mu D_{\gamma^{1/2}}\mu') \\ \Sigma_{12}^* &= \Sigma_{11} - \mu\nu^{-1}\Sigma_{12}' + \Sigma_{12}\nu'^{-1}\mu' - \mu\nu^{-1}\Sigma_{22}\nu'^{-1}\mu' \\ &= \Sigma_{11} - \mu D_{\gamma/12}\mu' + \mu D_{\gamma^{1/2}}\mu' - \Sigma_{11} = 0. \end{aligned} \tag{3.6}$$

This means that the transformed p-set $\mathbf{x}_1^*$ is uncorrelated with transformed p-set $\mathbf{x}_2^*$. We shall put simultaneous confidence bounds on the largest and smallest characteristic roots of $\lambda\lambda'$, i.e., of $\mu\nu^{-1}\nu'^{-1}\mu'$ and then show at the end of this section how these roots are, in a sense, a generalization of σ_1^2/σ_2^2 for case (i). We may note here, incidentally, that for $p = 1$, λ does, in fact, reduce to σ_1/σ_2. Next, denoting by S^* the sample dispersion matrix of $\mathbf{x}^*$, we have

$$(3.7)\qquad S^*(2p \times 2p) = \underset{p\qquad p}{\begin{bmatrix} S_{11}^* & S_{12}^* \\ S_{22}^{*\prime} & S_{22}^* \end{bmatrix}}\begin{matrix} p \\ p \end{matrix} \quad \text{(say)} \quad = ASA'$$
$$= \begin{bmatrix} I & -\lambda \\ I & \lambda \end{bmatrix}\begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix}\begin{bmatrix} I & I \\ -\lambda' & \lambda' \end{bmatrix},$$

whence we have

$$(3.8)\qquad \begin{aligned} S_{11}^* &= S_{11} - \lambda S_{12}' - S_{12}\lambda' + \lambda S_{22}\lambda', \\ S_{12}^* &= S_{11} - \lambda S_{12}' + S_{12}\lambda' - \lambda S_{22}\lambda', \\ S_{22}^* &= S_{11} + \lambda S_{12}' + S_{12}\lambda' + \lambda S_{22}\lambda'. \end{aligned}$$

Now we go back to (3.6). Note that, since $\Sigma_{12}^* = 0$, the transformed $\mathbf{x}_1^*$-set is uncorrelated with the transformed $\mathbf{x}_2^*$-set, and also that, in this case, the joint distribution of the canonical correlation coefficients and also, in particular, of the largest canonical correlation coefficient is known. Thus we can find a $c_\alpha(p, p, n-1) = c_\alpha$ (say) such that

$$(3.9)\qquad P[c_{\max}(S_{11}^{*-1}S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime}) \leqq c_\alpha \mid \Sigma_{12}^* = 0] = 1 - \alpha.$$

The set over which the probability statement (3.9) is made, namely,

$$c_{\max}(S_{11}^{*-1}S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime}) \leqq c_\alpha,$$

can be used as an acceptance region for the hypothesis that $\mu\nu^{-1}$ has a particular (matrix) value, and, in particular, that $\mu\nu^{-1} = I(p)$, or in other words, $\Sigma_{11} = \Sigma_{22}$. The problem now is to start from (3.9), use (3.8) and try to obtain confidence bounds on functions connected with $\lambda(=\mu\nu^{-1})$. For this we proceed as follows. Let c be a characteristic root of the matrix in (3.9). Then

$$(3.10)\qquad | cS_{11}^* - S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime} | = 0.$$

With $c = 1 - 4d$, this reduces to

$$(3.11)\qquad | dS_{11}^* - \tfrac{1}{4}S_{11}^* + \tfrac{1}{4}S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime} | = 0.$$

Now, using (3.8), we have

$$(3.12)\qquad \begin{aligned} -\tfrac{1}{4}S_{11}^* &= -S_{11} + \tfrac{1}{4}(S_{12}^* + S_{12}^{*\prime} + S_{22}^*) \\ &= -S_{11} + \tfrac{1}{4}(S_{12}^* + S_{22}^*)S_{22}^{*-1}(S_{12}^{*\prime} + S_{22}^*) - \tfrac{1}{4}S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime}. \end{aligned}$$

Hence

$$\left| dS_{11}^* - S_{11} + \left(\frac{S_{12}^* + S_{22}^*}{2}\right) S_{22}^{*-1} \left(\frac{S_{12}^{*\prime} + S_{22}^*}{2}\right) \right| = 0 \tag{3.13}$$

or

$$| dS_{11}^* - S_{11} + (S_{11} + S_{12}\lambda')S_{22}^{*-1}(S_{11} + \lambda S_{12}') | = 0.$$

Next, we recall that for a non-singular $M_4(q \times q)$ we have

$$\begin{vmatrix} M_1 & M_2 \\ M_3 & M_4 \end{vmatrix} \begin{matrix} p \\ q \end{matrix} = |M_4| \, |M_1 - M_2 M_4^{-1} M_3| \tag{3.14}$$
$$\begin{matrix} p & q \end{matrix}$$

and, using this, we observe that (3.13) is equivalent to

$$\begin{vmatrix} S_{11} - dS_{11}^* & S_{11} + S_{12}\lambda' \\ S_{11} + \lambda S_{12}' & S_{11} + \lambda S_{12}' + S_{12}\lambda' + \lambda S_{22}\lambda' \end{vmatrix} = 0, \tag{3.15}$$

that is,

$$\begin{vmatrix} S_{11} - dS_{11}^* & S_{12}\lambda' + dS_{11}^* \\ S_{11} + \lambda S_{12}' & S_{12}\lambda' + \lambda S_{22}\lambda' \end{vmatrix} = 0,$$

that is,

$$\begin{vmatrix} S_{11} - dS_{11}^* & S_{12}\lambda' + dS_{11}^* \\ \lambda S_{12}' + dS_{11}^* & \lambda S_{22}\lambda' - dS_{11}^* \end{vmatrix} = 0,$$

that is,

$$\left| \begin{bmatrix} S_{11} & S_{12}\lambda' \\ \lambda S_{12}' & \lambda S_{22}\lambda' \end{bmatrix} \begin{matrix} p \\ p \end{matrix} - d \begin{bmatrix} S_{11}^* & -S_{11}^* \\ -S_{11}^* & S_{11}^* \end{bmatrix} \right| = 0.$$
$$\begin{matrix} p & p \end{matrix}$$

But we have

$$\begin{bmatrix} S_{11} & S_{12}\lambda' \\ \lambda S_{12}' & \lambda S_{22}\lambda' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \lambda' \end{bmatrix} \tag{3.16}$$

and

$$\begin{bmatrix} S_{11}^* & -S_{11}^* \\ -S_{11}^* & S_{11}^* \end{bmatrix} = \begin{bmatrix} I \\ -I \end{bmatrix} S_{11}^* [I \quad -I]$$

$$= \begin{bmatrix} I \\ -I \end{bmatrix} [I \quad -\lambda] \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix} \begin{bmatrix} I \\ -\lambda' \end{bmatrix} [I \quad -I].$$

Hence (3.15) reduces to

$$\left|\begin{bmatrix} I & 0 \\ 0 & \lambda \end{bmatrix} S \begin{bmatrix} I & 0 \\ 0 & \lambda' \end{bmatrix} - d \begin{bmatrix} I \\ -I \end{bmatrix} [I \quad -\lambda]\, S \begin{bmatrix} I \\ -\lambda' \end{bmatrix} [I \quad -I] \right| = 0,$$

which is equivalent to

$$(3.17) \qquad \left| eS - \begin{bmatrix} I & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} I \\ -I \end{bmatrix} [I \quad -\lambda]\, S \begin{bmatrix} I \\ -\lambda' \end{bmatrix} [I \quad -I] \begin{bmatrix} I & 0 \\ 0 & \lambda'^{-1} \end{bmatrix} \right| = 0,$$

where $e = 1/d$, which again reduces to

$$(3.18) \qquad | eI(2p \times 2p) - S^{-1} \beta S \beta' | = 0,$$

where

$$(3.19) \qquad \beta(2p \times 2p) = \begin{bmatrix} I & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} I \\ -I \end{bmatrix} [I \quad -\lambda] = \begin{bmatrix} I & -\lambda \\ -\lambda^{-1} & I \end{bmatrix}.$$

Now we go back to (3.9), recall that $e = 1/d = 4/(1 - c)$, put $e_\alpha = 4/(1 - c_\alpha)$, observe that "$c_{\max} \leqq c_\alpha$" is equivalent to "$e_{\max} \leqq e_\alpha$," and hence that (3.9) is equivalent to

$$P[c_{\max}[S^{-1}\beta S \beta'] \leqq e_\alpha \mid \Sigma_{12}^* = 0] = 1 - \alpha,$$

or to

$$(3.20) \qquad P\left[\frac{(\mathbf{a}'\beta\mathbf{b})^2}{(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b})} \leqq e_\alpha \frac{\mathbf{a}'S\mathbf{a}}{\mathbf{a}'\mathbf{a}} \cdot \frac{\mathbf{b}'S^{-1}\mathbf{b}}{\mathbf{b}'\mathbf{b}} \,\middle|\, \Sigma_{12}^* \, 0 \quad \text{for all non null} \quad \mathbf{a}(2p \times 1) \quad \text{and} \quad \mathbf{b}(2p \times 1) \right] = 1 - \alpha.$$

Next, consider, for all non null **a** and **b**, the statement

$$(3.21) \qquad \frac{(\mathbf{a}'\beta\mathbf{b})^2}{(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b})} \leqq e_\alpha \frac{\mathbf{a}'S\mathbf{a}}{\mathbf{a}'\mathbf{a}} \cdot \frac{\mathbf{b}'S^{-1}\mathbf{b}}{\mathbf{b}'\mathbf{b}}.$$

Now specialize $\mathbf{a}'(2p \times 1)$ and $\mathbf{b}'(2p \times 1)$ into $\underset{p\quad p}{[\mathbf{a}_1' \quad 0]}\, 1$ and $\underset{p\quad p}{[\mathbf{b}_1' \quad 0]}\, 1$, and also into $\underset{p\quad p}{[0 \quad \mathbf{a}_2']}\, 1$ and $\underset{p\quad p}{[0 \quad \mathbf{b}_2']}\, 1$.

We next set

$$(3.22) \qquad S^{-1}(2p \times 2p) = \underset{\quad p \qquad p}{\begin{bmatrix} S^{11} & S^{12} \\ S^{12'} & S^{22} \end{bmatrix}} \begin{matrix} p \\ p \end{matrix},$$

whence we have

$$(3.23) \qquad \begin{aligned} S^{11} &= (S_{11} - S_{12}\, S_{22}^{-1} S_{12}')^{-1}, \qquad S^{22} = (S_{22} - S_{12}'\, S_{11}^{-1} S_{12})^{-1}, \\ S^{12} &= -S^{11} S_{12}\, S_{22}^{-1} = -S_{11}^{-1} S_{12}\, S^{22}. \end{aligned}$$

Back in (3.21) we now observe that (3.21) implies

$$(3.24)\qquad \frac{(\mathbf{a}_1'\lambda\mathbf{b}_2)^2}{(\mathbf{a}_1'\mathbf{a}_1)(\mathbf{b}_2'\mathbf{b}_2)} \leqq e_\alpha \frac{\mathbf{a}_1' S_{11}\mathbf{a}_1}{\mathbf{a}_1'\mathbf{a}_1}\frac{\mathbf{b}_2' S^{22}\mathbf{b}_2}{\mathbf{b}_2'\mathbf{b}_2}$$

for all non null $\mathbf{a}_1$ and $\mathbf{b}_2$, and that (3.21) also implies

$$(3.25)\qquad \frac{(\mathbf{a}_2\lambda^{-1}\mathbf{b}_1)^2}{(\mathbf{a}_2'\mathbf{a}_2)(\mathbf{b}_1'\mathbf{b}_1)} \leqq e_\alpha \frac{\mathbf{a}_2' S_{22}\mathbf{a}_2}{\mathbf{a}_2'\mathbf{a}_2}\cdot\frac{\mathbf{b}_1' S^{11}\mathbf{b}_1}{\mathbf{b}_1'\mathbf{b}_1},$$

for all non null $\mathbf{a}_2$ and $\mathbf{b}_1$. If now we consider the left side of (3.24), then it follows from Cauchy's inequality that for all non null $\mathbf{b}_2$, $(\mathbf{a}_1'\lambda\mathbf{b}_2)^2/(\mathbf{a}_1'\mathbf{a}_1)(\mathbf{b}_2'\mathbf{b}_2) \leqq (\mathbf{a}_1'\lambda\lambda'\mathbf{a}_1)/(\mathbf{a}_1'\mathbf{a}_1)$, and it is also well known that for all non null $\mathbf{a}_1$, $c_{\min}(\lambda\lambda') \leqq (\mathbf{a}_1'\lambda\lambda'\mathbf{a}_1)/(\mathbf{a}_1'\mathbf{a}_1) \leqq c_{\max}(\lambda\lambda')$. We have also exactly similar results by interchanging $\mathbf{a}_1$ and $\mathbf{b}_2$, and similar results on the left side of (3.25), in terms of λ^{-1} and $\mathbf{a}_2$ and $\mathbf{b}_1$ and then again by the interchange of $\mathbf{a}_2$ and $\mathbf{b}_1$.

Next, maximizing the left side of (3.24) w.r.t. $\mathbf{a}_1$ and $\mathbf{b}_2$, we observe ([2], [3], [4]) that (3.24) and hence (3.21) $\Rightarrow$

$$c_{\max}(\lambda\lambda') \leqq e_\alpha c_{\max}(S_{11})c_{\max}(S^{22}),$$

or, after substitution from (3.23),

$$(3.26)\qquad c_{\max}(\lambda\lambda') \leqq e_\alpha c_{\max}(S_{11})/c_{\min}(S_{22} - S_{12}' S_{11}^{-1} S_{12}).$$

Likewise, maximizing the left side of (3.25) w.r t. $\mathbf{a}_2$ and $\mathbf{b}_1$, we observe [4] that (3.25) and hence (3.21) imply

$$(3\ 27)\qquad c_{\max}(\lambda'^{-1}\lambda^{-1}) \leqq e_\alpha c_{\max}(S_{22})c_{\max}(S^{11}).$$

Now recall that [3], since all non zero roots of $\lambda^{-1}\lambda'^{-1}$ are also roots of $\lambda'^{-1}\lambda^{-1}$, i.e., of $(\lambda\lambda')^{-1}$ and λ is nonsingular, therefore, $c_{\min}(\lambda^{-1}\lambda'^{-1}) = c_{\min}(\lambda\lambda')^{-1} = 1/c_{\max}(\lambda\lambda')$ and also similarly that $c_{\min}(\lambda^{-1}\lambda'^{-1}) = 1/c_{\max}(\lambda\lambda')$. At this point, using (3.23) we observe that (3.27) and hence (3.25) and hence (3.21) imply

$$(3.28)\qquad c_{\min}(\lambda\lambda') \geqq \frac{1}{e_\alpha} c_{\min}(S_{11} - S_{12} S_{22}^{-1} S_{12}') \Big/ c_{\max}(S_{22}).$$

Also, going back to (3.24) and first maximizing the left side of it w.r.t $\mathbf{b}_2$ and then minimizing the right side w.r.t. $\mathbf{a}_1$, we observe [4] that (3.24) and hence (3.21) imply

$$(3.29)\qquad c_{\min}(\lambda\lambda') \leqq e_\alpha c_{\min}(S_{11})/c_{\min}(S_{22} - S_{12}' S_{11}^{-1} S_{12}),$$

and, furthermore, first maximizing the left side w.r.t. $\mathbf{a}_1$ and then minimizing the right side w.r.t. $\mathbf{b}_2$, we observe [4] that (3.24) and hence (3.21) also imply

$$(3.30)\qquad c_{\min}(\lambda\lambda') \leqq e_\alpha c_{\max}(S_{11})/c_{\max}(S_{22} - S_{12} S_{22}^{-1} S_{12}').$$

Likewise, back in (3.25), first maximizing the left side w.r.t. $\mathbf{b}_1$ and then minimizing the right side w.r.t. $\mathbf{a}_2$, we observe [4] that (3.25) and hence (3.21) imply

$$(3.31) \qquad c_{\max}(\lambda\lambda') \geqq \frac{1}{e_\alpha} c_{\min}(S_{11} - S_{12} S_{22}^{-1} S_{12}') \Big/ c_{\min}(S_{22}),$$

and first maximizing the left side w.r.t. $\mathbf{a}_2$ and then minimizing the right side w.r.t. $\mathbf{b}_1$, we observe [4] that (3.25) and hence (3.21) also imply

$$(3.32) \qquad c_{\max}(\lambda\lambda'(\geqq \frac{1}{e_\alpha} c_{\max}(S_{11} - S_{12} S_{22}^{-1} S_{12}') \Big/ c_{\max}(S_{22}).$$

Now combining (3.26), (3.28), (3.29)–(3.32), we observe that (3.21) implies all these statements, and hence, going back to (3.20), we have with a joint probability $\geqq 1 - \alpha$, the bounds

$$(3.33) \qquad \frac{1}{e_\alpha} c_{\min}(S_{11} - S_{12} S_{22}^{-1} S_{12}') \Big/ c_{\max}(S_{22}) \leqq c_{\min}(\lambda\lambda')$$
$$\leqq e_\alpha \min [c_{\min}(S_{11})/c_{\min}(S_{22} - S_{12}' S_{11}^{-1} S_{12}), c_{\max}(S_{11})/c_{\max}(S_{22} - S_{12}' S_{22}^{-1} S_{12})]$$

and

$$(3.34) \qquad \frac{1}{e_\alpha} \max [c_{\min}(S_{11} - S_{12} S_{22}^{-1} S_{12}')/c_{\min}(S_{22}), c_{\max}(S_{11} - S_{12} S_{22}^{-1} S_{12}')/c_{\max}(S_{22})]$$
$$\leqq c_{\max}(\lambda\lambda') \leqq e_\alpha c_{\max}(S_{11})/c_{\min}(S_{22} - S_{12}' S_{11}^{-1} S_{12}).$$

It is interesting to use [3] and check that the lower bound of (3.33) is $\leqq$ the upper bound of (3.34), but that the upper bound of (3.33) might be $\geqq$ or $<$ the lower bound of (3.34). However, it is to be always remembered that $c_{\min}(\lambda\lambda') \leqq c_{\max}(\lambda\lambda')$, which should imply an obvious restriction on combined bounds on $c_{\max}(\lambda\lambda')$ and $c_{\min}(\lambda\lambda')$.

Truncation. Going back to (3.24) again we can proceed as in [4], equate to zero any element of $\mathbf{a}_1$ and the *corresponding* elements of $\mathbf{b}_2$, $\mathbf{a}_2$, and $\mathbf{b}_1$ (it has to be the corresponding elements, in order to make the process physically meaningful) and then apply the process of maximization, minimization, etc., leading ultimately to the same kind of statements as (3.33) and (3.34) in terms, however, of truncated matrices everywhere, with one variate of the first p-set and the corresponding variate of the second p-set being cut out. Thus there will be $\binom{p}{1}$, i.e., p pairs of such statements. Likewise equating to zero any two elements of $\mathbf{a}_1$ and the corresponding elements of $\mathbf{b}_2$, $\mathbf{a}_2$ and $\mathbf{b}_1$, we are ultimately led to $\binom{p}{2}$, pairs of statements like (3.33) and (3.34) based on different possible sets of $(p - 2)$ variates, and so on. Ultimately we have $1 + \binom{p}{1} + \binom{p}{2} + \cdots + \binom{p}{p-1}$, i.e., $2^p - 1$ pairs of statements like (and including) (3.33) and (3.34) with a joint probability $\geqq 1 - \alpha$. It should be noticed that on all these statements e_α, however, stays the same.

It follows from the above remarks that, with a joint confidence coefficient $\geqq 1 - \alpha$, (3.33) and (3.36) imply, among other things, the following set of confidence statements on the ratios $\sigma_{1i}^2/\sigma_{2i}^2$:

$$(3.34.1) \qquad \frac{1}{c_\alpha}\frac{s_{1i}^2}{s_{2i}^2}(1 - r_i^2) \leqq \frac{\sigma_{1i}^2}{\sigma_{2i}^2} \leqq c_\alpha \frac{s_{1i}^2}{s_{2i}^2(1 - r_i^2)} \qquad \text{for } i = 1, 2, \cdots, p,$$

where s_{1i}^2, s_{2i}^2, σ_{1i}^2, σ_{2i}^2 and r_i stand respectively for the sample variances of the ith variate for the two sets, the population variances of the ith variate for the two sets and the sample correlation coefficient between the ith variate for the first set and for the second set.

Interpretation of the role of the characteristic roots of $\lambda\lambda'$. The characteristic roots of $\lambda\lambda'$, i.e., of $\mu\nu^{-1}\nu'^{-1}\mu'$ are all equal to unity if and only if $\mu\nu^{-1}\nu'^{-1}\mu'$ is an identity matrix, i.e., if and only if

$$(3.35) \qquad \mu\nu^{-1} = A, \quad \text{i.e.,} \quad \mu = A\nu,$$

where A is any arbitrary orthogonal matrix. Going back to (3.2) we easily check that (3.35) implies

$$(3.36) \qquad \Sigma_{11} = A\Sigma_{22}A',$$

which, if we recall that A is orthogonal, and Σ_{11} and Σ_{22} are symmetric, is precisely the condition that Σ_{11} and Σ_{22} are to be similar matrices. Furthermore, using (3.2) again it is easy to see that (3.35) also implies

$$(3.37) \qquad \Sigma_{12} = \mu D_{\gamma^{1/2}}\nu' = A\nu D_{\gamma^{1/2}}\nu' = A \times \text{a symmetric matrix},$$

where A is the same orthogonal matrix that occurs in (3.36). Thus (3.35) implies (3.36) and (3.37) and it is also easy to verify that (3.36) and (3.37) imply (3.35). Hence all the characteristic roots of $\lambda\lambda'$, i.e., of $\mu\nu^{-1}\nu'^{-1}\mu'$ being unity is a necessary and sufficient condition that the relations (3.36) and (3.37) should hold. The deviation of these characteristic roots from unity might be regarded as a (joint) measure of departure from the hypothesis given by (3.36) and hence (3.37), of which a very special case is the one that we get for the [illegible] problem. Further statistical implications of (3.36) and (3.37) will be discussed in a later paper.

4. Confidence bounds for the case (iii). Starting from the [illegible] distribution characterized in section 2, put $q = $ [illegible] variate $z = x_1 - qx_2$ (assume that $\xi_2 \neq 0$, i.e., $q \neq$ [illegible] where $\sigma_z^2 = \sigma_1^2 - 2q\rho\sigma_1\sigma_2 + q^2\sigma_2^2$. Thus

$$\sqrt{n}\,\bar{z}/s_z = \sqrt{n}(\bar{x}_1 - q\bar{x}_2)/(s_1^2 - \text{[illegible]}$$

has the (central) t-distribution with d.f. $(n - 1)$, [illegible] that

$$P\left[n(\bar{x}_1 - q\bar{x}_2)^2/(s_1^2 - 2qs_1s_2r + \text{[illegible]}\right.$$

or

$$(4.1)\quad P[(\bar{x}_2^2 - ks_2^2)q^2 - 2(\bar{x}_1\bar{x}_2 - ks_1s_2r)q + (\bar{x}_1^2 - ks_1^2) \leqq 0] = 1 - \alpha,$$

where $k = (1/n)t_{\alpha/2}^2$. We can use the statement within the parentheses in (4.1) as an acceptance region for the hypothesis that the population ratio of means has a specific value q. But such an acceptance is, of course, well known, at least in an implicit form.

Subject to the restriction that q is to have real values, the statement within the parentheses in (4.1) gives the confidence bounds on $q = \xi_1/\xi_2$. There is also the further restriction that (4.1) is supposed to be a probability statement on $\bar{x}_1$, $\bar{x}_2$, s_1 and s_2 for all real values of $q = \xi_1/\xi_2$, except for $\xi_2 = 0$, i.e., for $q = \pm\infty$. Equating to zero the expression on the left side of the inequality statement under the probability sign in (4.1), we have an equation in q whose coefficients involve stochastic variates. The actual confidence bounds on q are given by

$$(4.2)\quad \begin{aligned} &\frac{(\bar{x}_1\bar{x}_2 - ks_1s_2r) - [(\bar{x}_1\bar{x}_2 - ks_1s_2r)^2 - (\bar{x}_1^2 - ks_1^2)(\bar{x}_2^2 - ks_2^2)]^{\frac{1}{2}}}{(\bar{x}_2^2 - ks_2^2)} \leqq q \\ &\leqq \frac{(\bar{x}_1\bar{x}_2 - ks_1s_2r) + [(\bar{x}_1\bar{x}_2 - ks_1s_2r)^2 - (\bar{x}_1^2 - ks_1^2)(\bar{x}_2^2 - ks_2^2)]^{\frac{1}{2}}}{(\bar{x}_2^2 - ks_2^2)}. \end{aligned}$$

The bounds will be physically meaningful only if the expression under the radical is non-negative, i.e., only if,

$$(4.3)\quad \frac{\bar{x}_1^2}{s_1^2} + \frac{\bar{x}_2^2}{s_2^2} \geqq 2\frac{\bar{x}_1}{s_1}\cdot\frac{\bar{x}_2}{s_2}r + k\cdot\frac{\bar{x}_1^2}{s_1^2}\cdot\frac{\bar{x}_2^2}{s_2^2}(1 - r^2).$$

Notice that $(\bar{x}_1^2/s_1^2) + (\bar{x}_2^2/s_2^2)$ is always greater than or equal to $2(\bar{x}_1/s_1)(\bar{x}_2/s_2)r$ but may not always be greater than or equal to the right side of (4.3). Thus, if in the sample, the inequality (4.3) breaks down we should not, in that situation, attempt to put any confidence bounds on ξ_1/ξ_2.

Going back to (4.1) and tying it up with (4.2) and (4.3) we now observe that α is the probability of choosing a sample such that either (4.2) is not a real interval or (4.2) is real but does not cover the true value.

5. Confidence bounds for the case (iv). Starting from the $(p + p)$ variate normal distribution characterized in section 3, define a set of q's, $q_1, q_2, \cdots, q_p$ by $\xi_1 = D_q\xi_2$ where $D_q(p \times p)$ is a diagonal matrix whose diagonal elements are $q_1, \cdots, q_p$. Introduce a new variate $\mathbf{z}(p \times 1)$ defined by

$$(5.1)\quad \mathbf{z}(p \times 1) = \underset{p}{p[I} \quad \underset{p}{-D_q]} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \begin{matrix} p \\ p \end{matrix} = A(p \times 2p)\mathbf{x}(2p \times 1) \quad \text{(say)}.$$

It is easy to check that $E(\mathbf{y}) = \xi_1 - D_q\xi_2 = 0$, whence $\mathbf{z}$ is $N(0, \Sigma_z)$ where $\Sigma_z = A\Sigma A'$. Also, given the sample dispersion matrix of $\mathbf{x}(2p \times 1)$, in the form

$$(5.2)\quad S(2p \times 2p) = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22}, \end{bmatrix}$$

we have sample dispersion matrix of $\mathbf{z}(p \times 1)$ given by

$$S_z = ASA' = S_{11} - D_q S_{22}' - S_{12} D_q + D_q S_{22} D_q . \tag{5.3}$$

Also the sample mean vector of $\mathbf{z}(p \times 1)$ is given by

$$\bar{\mathbf{z}} = \bar{\mathbf{x}}_1 - D_q \bar{\mathbf{x}}_2 . \tag{5.4}$$

Thus, with the q's defined as above, $n\bar{\mathbf{z}}' S_z^{-1} \bar{\mathbf{z}}$ is distributed as (central) Hotelling's T^2, which means that we can find a T_α^2 such that

$$P\left[\bar{\mathbf{z}}' S_z^{-1} \bar{\mathbf{z}} \leqq \frac{1}{n} T_\alpha^2 \mid q\text{'s defined as above}\right] = 1 - \alpha. \tag{5.5}$$

The set over which the probability statement (5.5) is made, can be used as an acceptance region for the hypothesis that the population mean ratios have specific values q_i's. This, of course, is implicit in the possible applications of Hotelling's T^2. Now consider the statement within the parentheses in (5.5) It is well known that this statement is equivalent to the statement that all $c[\bar{\mathbf{z}}\bar{\mathbf{z}}' S_z^{-1}] \leqq T_\alpha^2/n$, which again is equivalent to

$$\frac{\mathbf{a}'\bar{\mathbf{z}}\bar{\mathbf{z}}'\mathbf{a}}{\mathbf{a}'\mathbf{a}} \leqq \frac{T_\alpha^2}{n} \cdot \frac{\mathbf{a}' S_z \mathbf{a}}{\mathbf{a}'\mathbf{a}}, \tag{5.6}$$

for all non null $\mathbf{a}(p \times 1)$'s. Considering the left side of (5.6), we use again Cauchy's inequality to obtain that for all non null $\mathbf{a}$'s, $\mathbf{a}'\bar{\mathbf{z}}/(\mathbf{a}'\mathbf{a})^{\frac{1}{2}} \leqq +(\bar{\mathbf{z}}'\bar{\mathbf{z}})^{\frac{1}{2}}$ whence we see that under variation of $\mathbf{a}$ the largest value of the left side of (5.6) $= \bar{\mathbf{z}}'\bar{\mathbf{z}}$, that is, $= \sum_{i=1}^{p} (\bar{x}_{1i} - q_i \bar{x}_{2i})^2$, where $\bar{x}_{1i}$ and $\bar{x}_{2i}$ (for $i = 1, 2, \cdots, p$) stand for the ith elements of the vectors $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$. We also note that, aside from the constant factor T_α^2/n, the largest value of the right side of (5 6) under variation of $\mathbf{a}$'s is $c_{\max}(S_z)$, i.e., $c_{\max}(ASA')$, i.e., $c_{\max}(SA'A)$. Now we use [1] to obtain that

$$\begin{aligned} c_{\max}(SA'A) &\leqq c_{\max}(S) c_{\max}(A'A), \text{ i.e., } \leqq c_{\max}(S) c_{\max}(AA'), \\ &\text{i.e., } \leqq c_{\max}(S) c_{\max}[I + D_{q^2}], \\ &\text{i.e., } \leqq c_{\max}(S) \max[1 + q_1^2, 1 + q_2^2, \cdots, 1 + q_p^2]. \end{aligned} \tag{5.7}$$

Now, if we go back to (5.6) and maximize the left side w.r.t. $\mathbf{a}$, it is easy to check that (5.6) implies

$$\frac{1}{n} T_\alpha^2 c_{\max}(S) \max[1 + q_1^2, 1 + q_2^2, \cdots, 1 + q_p^2] - \sum_{i=1}^{p} (\bar{x}_{1i} - q_i \bar{x}_{2i})^2 \geqq 0$$

or

$$\frac{1}{n} T_\alpha^2 c_{\max}(S) \max[1 + q_1^2, \cdots, 1 + q_p^2] - \bar{\mathbf{x}}_1'\bar{\mathbf{x}}_1 - \sum_{i=1}^{p} q_i^2 \bar{x}_{2i}^2 + 2 \sum_{i=1}^{p} q_i \bar{x}_{1i} \bar{x}_{2i} \geqq 0. \tag{5.8}$$

Also notice that

$$(5.9)\qquad \left|\sum_{i=1}^{p} q_1\, \bar{x}_{1i}\, \bar{x}_{2i}\right| \leqq \sum_{i=1}^{p} \left| q_i \right| \left| \bar{x}_{1i}\, \bar{x}_{2i} \right| \leqq [\max(q_1^2, \cdots, q_p^2)]^{\frac{1}{2}} \sum_{i=1}^{p} | \bar{x}_{1i}\, \bar{x}_{2i} |,$$

and

$$- \sum_{i=1}^{p} q_i^2\, \bar{x}_{2i}^2 \leqq - \min(q_1^2, \cdots, q_p^2) \sum_{i=1}^{p} \bar{x}_{2i}^2 .$$

Hence it is easy to check that (5.8) and hence (5.6) imply

$$(5.10)\qquad \frac{1}{n} T_\alpha^2\, c_{\max}(S) \max[1 + q_1^2, \cdots, 1 + q_p^2] + 2 \sum_{i=1}^{p} | \bar{x}_{1i}\, \bar{x}_{2i} | \max[q_1^2, \cdots, q_p^2]^{\frac{1}{2}} - \bar{\mathbf{x}}_1' \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2' \bar{\mathbf{x}}_2 \min(q_1^2, \cdots, q_p^2) \geqq 0.$$

Going back to (5.5) we now observe that with a probability $\geqq 1 - \alpha$, we have the confidence statement (5.8) or (5.10).

Truncation. Here again, as in section 4, it is possible to go back to (5.6), proceed in the same way as before and get statements like (5.8) or (5.10) on any $(p - 1)$ variate-pairs, or on any $(p - 2)$ variate-pairs, and so on, and finally any variate-pair, thus ultimately obtaining $2^p - 1$ confidence statements like (5.8) or (5.10), all of them with a joint confidence coefficient $> 1 - \alpha$.

If we are interested in pairwise comparisons we go back to (5.6), set $k = T_2^2/n$ and choose $\mathbf{a}$ to be the vector with 1 in the ith position and 0's elsewhere. The resulting inequality can be written as (4.2) (with $k = T_\alpha^2/n$). Thus (5.6) implies a set of inequalities like this for $i = 1, 2, \cdots, p$, and hence, with a confidence coefficient greater than or equal to a preassigned $1 - \alpha$, we have the set of confidence bounds on ξ_{1i}/ξ_{2i} given by

$$(5.11)\qquad (e_{1i} - e_{2i}^{\frac{1}{2}})/e_{3i} \leqq q_i = \xi_{1i}/\xi_{2i} \leqq (e_{1i} + e_{2i}^{\frac{1}{2}})/e_{3i},$$

where, for $i = 1, 2, \cdots, p$,

$$(5.12)\qquad \begin{aligned} e_{1i} &= \bar{x}_{1i}\bar{x}_{2i} - k s_{1i} s_{2i} r_{12i}, \qquad e_{3i} = \bar{x}_{2i}^2 - k s_{2i}^2, \\ e_{2i} &= (\bar{x}_{1i}\bar{x}_{2i} - k s_{1i} s_{2i} r_{12i})^2 - (\bar{x}_{1i}^2 - k s_{1i}^2)(\bar{x}_{2i}^2 - k s_{2i}^2). \end{aligned}$$

As in section 4, the bounds will be physically meaningful only if

$$(5.13)\qquad \frac{x_{1i}^2}{s_{1i}^2} + \frac{\bar{x}_{2i}^2}{s_{2i}^2} \geqq 2\, \frac{\bar{x}_{1i}\, \bar{x}_{2i}}{s_{1i}\, s_{2i}}\, r_{12i} + k\, \frac{\bar{x}_{1i}^2\, \bar{x}_{2i}^2}{s_{1i}^2\, s_{2i}^2} (1 - r_{12i}^2).$$

As in section 4 so also here, the remarks made after (4.3) will be pertinent again as an indication of how to use these bounds.

In conclusion it is a great pleasure to thank the referee and the associate editor for their valuable comments and suggestions. The result (5.11), in particular, is entirely due to the referee and provides shorter bounds than the ones obtained by the authors' originally, starting from (5.10) rather than directly from (5.6).

REFERENCES

[1] S. N. Roy, "A useful theorem in matrix theory," *Proc. Amer. Math. Soc.*, Vol 5 (1954), pp. 635–638.
[2] S. N. Roy, "Some further results in simultaneous confidence interval estimation," *Ann Math. Stat.*, Vol. 25 (1954), pp. 752–761.
[3] S. N. Roy, "A report on some aspects of multivariate analysis," Institute of Statistics mimeograph series no. 121.
[4] S N. Roy and R. Gnanadesikan, "Further contributions to multivariate confidence bounds," Institute of Statistics mimeograph series no. 1551. (Also *Biometrika*, Vol. 44 (1957), pp. 399-410.
[5] Helen Walker and Joseph Lev, "Statistical Inference," Henry Holt and Company, New York, 1953.

A THREE-SAMPLE KOLMOGOROV-SMIRNOV TEST

By Herbert T. David

Iowa State College

1. Introduction. In 1951, Gnedenko and Korolyuk published an elegant derivation ([6])[1] of the null distribution of the Kolmogorov-Smirnov statistic $D_{2,n}$ for two samples of equal size n. The statistic $D_{2,n}$ is given by

$$D_{2,n} = \sup_t | F_{2,n}(t) - F_{1,n}(t) | , \tag{1}$$

where $F_{i,n}(t)$ is the sample cumulative distribution function for the ith sample. The distribution derived by Gnedenko and Korolyuk is

$$\Pr\left\{D_{2,n} \geqq \frac{l}{n}\right\} = 2\binom{2n}{n}^{-1} \sum_{i=1}^{[n/l]} (-1)^{i+1} \binom{2n}{n-il}. \tag{2}$$

Since

$$\lim_{n\to\infty} \frac{\binom{2n}{n-k\sqrt{n}}}{\binom{2n}{n}} = e^{-k^2}, \tag{3}$$

(2) easily leads to the familiar asymptotic result

$$\lim_{n\to\infty} \Pr\left\{n^{1/2}D_{2,n} \geqq \lambda\right\} = 2\sum_{i=1}^{\infty} (-1)^{i+1} e^{-(i\lambda)^2}. \tag{4}$$

Gnedenko and Korolyuk's proof hinges on the fact that, in the null case (for two samples drawn from the same continuous distribution), $\Pr\{D_{2,n} \geqq l/n\}$ equals the probability that the maximum deviation from the origin of a certain random walk in the line is at least l. The random paths involved in this random walk start at the origin, and consist of $2n$ unit steps, n to the left and n to the right, with all possible permutations of left and right steps equally likely. The probability $\Pr\{D_{2,n} \geqq l/n\}$ is thus equal to, say, $M \Big/ \binom{2n}{n}$, where $\binom{2n}{n}$ is the total number of equally likely paths, and M is the number of these paths with maximum deviation from the origin at least l. M can be computed by the reflection principle in the line ([2], [1]), leading to (2).

In this paper I show that the null distribution of the three-sample extension $D_{3,n}$ (see (6) below) of $D_{2,n}$ can be derived by extending the geometric approach of [6] from the line to the plane.

Received October 21, 1957; revised February 25, 1958.

[1] The review of this paper in Mathematical Reviews [3] was brought to my attention by Murray Rosenblatt.

$D_{3,n}$ is but one of several "distance" criteria that have recently appeared in the literature. Fisz and Kiefer [4], [7][2] have shown that the criterion

$$R_n = \max \{\sup_t | F_{3,n}(t) - F_{2,n}(t) |, \sup_t | 2F_{1,n}(t) - F_{2,n}(t) - F_{3,n}(t) | \},$$

and extensions of R_n to k samples and unequal sample sizes, can be used with existing Kolmogorov-Smirnov tables because the events

$$A: [\sup_t | F_{3,n}(t) - F_{2,n}(t) | \leqq \lambda_1]$$

and

$$B: [\sup_t | 2F_{1,n}(t) - F_{2,n}(t) - F_{3,n}(t) | \leqq \lambda_2]$$

are independent. It may be of interest to note that the criterion R_n corresponds to using a rectangular boundary on the hexagonal grid of Figure 1, and that the independence of the events A and B, and distribution of R_n, follow easily from this representation.

Ozols' [8][2] treatment of the criterion

$$S_n = \max \{\sup_t (F_{3,n}(t) - F_{2,n}(t)), \sup_t (F_{2,n}(t) - F_{1,n}(t))\},$$

is similar to my treatment of the criterion $D_{3,n}$. The boundary corresponding to S_n is an infinite 60° wedge on the hexagonal grid of Figure 1.

Finally, Kiefer [7] and Gihman [5][3] consider a criterion T_n (or D_k^2) of form

$$\sup_t \left(\sum_{i=1}^{k} (F_{i,n}(t) - \overline{F_n(t)})^2 \right), \qquad \overline{F_n(t)} = \sum_{i=1}^{k} F_{i,n}(t)/k,$$

and extensions of this criterion to unequal sample sizes; Kiefer [7] also considers the k-sample extension V_n of the statistic (5) given below in section 2.

Kiefer has shown in [7] that "distance" criteria of the type discussed above have good power properties. Among such criteria, one might suspect on heuristic grounds that $D_{3,n}$ has especially good power characteristics against the "one-sided" alternative H_A: $[(X < Y < Z)$ or $(Y < Z < X)$ or $(Z < X < Y)]$. This is because H_A tends to generate paths, on the grid of Figure 1, in the directions $\pi/6$, $\pi/6$, $+2\pi/3$, or $\pi/6 + 4\pi/3$.

2. A three-sample Kolmogorov-Smirnov statistic and its small-sample null distribution. A natural three-sample extension of (1) would be

$$\text{Max} \{\sup_t | F_{2,n}(t) - F_{1,n}(t) |, \sup_t | F_{3,n}(t) - F_{2,n}(t) |, \sup_t | F_{1,n}(t) - F_{3,n}(t) | \}. \tag{5}$$

[2] I owe these references to an associate editor.
[3] I owe this reference to Milton Sobel

But (5) does not lend itself easily to an extension of Gnedenko and Korolyuk's geometric method; a statistic that does so lend itself is that obtained from (5) by deleting the absolute value signs:

$$(6)\qquad D_{3,n} = \text{Max}\,\{\sup_t\,(F_{2,n}(t) - F_{1,n}(t)), \sup_t\,(F_{3,n}(t) - F_{2,n}(t)), \sup_t\,(F_{1,n}(t) - F_{3,n}(t))\}.$$

The null distribution of $D_{3,n}$ is its distribution when the three samples are drawn from the same continuous population. This null distribution is derived as follows.

A step of type A in the plane is defined to be a unit step to the right (direction 0); a step of type B is a unit step in the direction $2\pi/3$, and a step of type C is a unit step in the direction $4\pi/3$.

In the null case considered, ties occur with probability zero; hence (almost) every set of three samples of n leads to a ranking of the $3n$ sample values making up the three samples. Corresponding to each set of three samples, consider a path $p_{3,n}$ from the origin, composed of $3n$ unit steps, with the kth step of $p_{3,n}$ of type A if the rank k belongs to the first sample, etc. Clearly every $p_{3,n}$ contains n steps of each of the three types A, B and C.

Next, consider the equilateral triangle in the plane that is centered at the origin, has sides of length $3l$, and is oriented such that one of its sides is horizontal. Call this equilateral triangle Γ_l. Clearly

$$(7)\qquad \left\{D_{3,n} \geqq \frac{l}{n}\right\} \Leftrightarrow \{(p_{3,n} \cap \Gamma_l) \text{ is not empty}\}.$$

But in the null case every path $p_{3,n}$ (permutation of $3n$ steps, n each of type A, B and C) is possible, and each of the $(3n)!/(n!)^3$ such paths is equally likely. Hence (7) implies

$$(8)\qquad \Pr\left\{D_{3,n} \geqq \frac{l}{n}\right\} = \Pr\,\{(P_{3,n} \cap \Gamma_l) \text{ is not empty}\} = N/(3n)!/(n!)^3,$$

where N is the number of paths $p_{3,n}$ touching or piercing Γ_l. The small-sample problem is therefore solved if N can be evaluated.

N is evaluated by extending to the plane the principle of reflection that yielded M. Consider a hexagonal grid in the plane, consisting of "$\oplus$" points and "$\ominus$" points, as indicated in figure 1 for the case ($n = 7$, $l = 2$). The extent of the grid is fixed by the fact that the distance between the origin 0 and each of the three "vertices" (indicated by the letters V_1, V_2, V_3 in figure 1) is $(3l)([n/l])$. This distance is of course $(3.2)([7/2]) = 18$ for the case illustrated by figure 1. The central triangle indicated by the heavy line in figure 1 represents Γ_l.

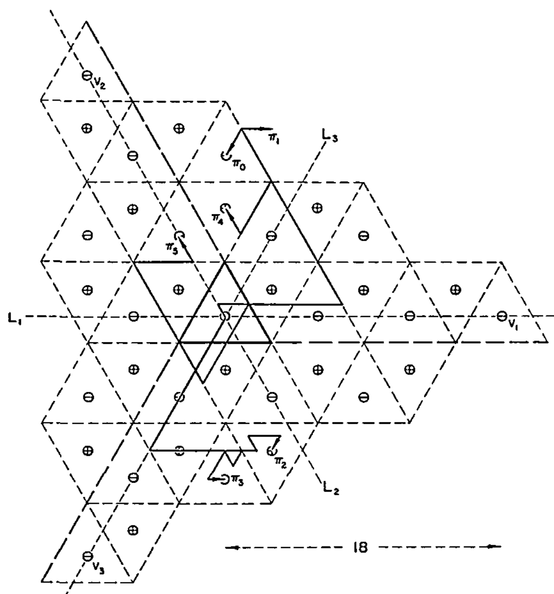


Fig. 1

Any path from the origin 0 to a ⊕ point, that consists of $3n$ steps of type A, B or C, is called a path of type ⊕. A path of type ⊖ is defined similarly. A path of type ⊕ or ⊖ is called an auxiliary path π. Again, any path from the origin to the origin, that consists of $3n$ steps of type A, B or C, and that touches the boundary Γ_t, is called a boundary path β. Finally, $N_\oplus$, $N_\ominus$ and N_0 are, respectively, the total number of paths of type ⊕, the total number of paths of type ⊖, and the total number of boundary paths.

The argument now is as follows.

(9) For any particular endpoint, whether it be a ⊕ point, a ⊖ point, or the origin 0, the specification that there be $3n$ steps in a boundary path or auxiliary path from the origin to that endpoint actually determines the numbers m_A, m_B and m_C of steps of types A, B and C involved in the path.

(9) follows from the fact that the location of the endpoint provides three equations in m_A, m_B and m_C, which, together with

$$(10a) \qquad m_A + m_B + m_C = 3n,$$

yield m_A, m_B and m_C. These three equations are

$$(10b) \qquad m_C - m_B = K_1$$

$$(10c) \qquad m_A - m_C = K_2$$

$$(10d) \qquad m_B - m_A = K_3.$$

K_1, K_2 and K_3 are determined by the signed perpendicular distances d_1, d_2 and d_3 of the endpoint from the lines L_1, L_2 and L_3 (see figure 1). For example, $K_1 = 2d_1/3^{\frac{1}{2}}$ if the endpoint is d_1 units below L_1, and $K_1 = -2d_1/3^{\frac{1}{2}}$ if the endpoint is d_1 units above L_1.

In particular, for every path from the origin to the origin, (10b), (10c) and (10d) become $m_C - m_B = m_A - m_C = m_B - m_A = 0$, which, together with (10a), yield $m_A = m_B = m_C = n$. This last implies that the boundary paths are exactly the paths enumerated by N, or

$$(11) \qquad N_0 = N.$$

Next, we introduce the operation of reflection. Reflection is an operation performed on an auxiliary path π that yields a path $p(\pi)$ which can be either an auxiliary path or a boundary path. Reflection is defined as follows.

Let π be an auxiliary path whose last point of contact (proceeding along π from the origin) with Γ_l is the point u.

1) Suppose first that u is not a vertex of Γ_l. Suppose for example that u lies within the horizontal side of Γ_l (i.e. the side oriented in the direction of a step of type A.) Then $p(\pi)$ is obtained from π by replacing every step of type B occurring after u by a step of type C, and every step of type C occurring after u by a step of type B. Analogously, if u lies within the side of Γ_l oriented in the direction of a step of type B, then $p(\pi)$ is obtained from π by replacing steps of type A occurring after u by steps of type C, and vice-versa; if u lies within the side of Γ_l oriented in the direction of a step of type C, the transposition of step types involves types A and B.

For example, reflection of the path π_4 (see figure 1) leads to the path π_5.

2) If u is a vertex of Γ_l, then reflection consists, as in 1), of a transposition of two step types. Which two step types are involved is determined by the requirement that the step occurring immediately after u be converted into a step lying in Γ_l. Thus, for example, if u is the vertex of Γ_l lying on both of the two non-horizontal sides of Γ_l, and if the step occurring immediately after u is a step of type A, then the two step types involved in the transposition are types A and C; in other words, $p(\pi)$ is obtained from π by replacing every step of type A occurring after u by a step of type C, and every step of type C occurring after u by a step of type A.

The operations of reflection, performed on an auxiliary path π, yields a path $p(\pi)$ which 1) contains $3n$ steps, 2) contains no steps of types other than A, B and C; and 3) begins at the origin and ends at a $\oplus$ point, at a $\ominus$ point, or at the origin. (Endpoints exterior to the grid of figure 1, such as the endpoint of the path π_1 for example, cannot result from reflection, because, for such endpoints, equations (10) have at least one negative solution). Finally, it is clear that: 4) the number of steps of π exterior to Γ_l, from the point of last contact of π with Γ_l to the endpoint of π, is greater by at least one than the number of steps of $p(\pi)$ exterior to Γ_l, from the point of last contact of $p(\pi)$ with Γ_l to the endpoint of $p(\pi)$.

By 1), 2) and 3), $p(\pi)$ is either an auxiliary path or a boundary path, and, by 4), successive reflection $p_1(\pi)$, $p_2(p_1(\pi))$, $p_3(p_2(p_1(\pi)))$, $\cdots$ eventually lead to a boundary path, say $p_k(p_{k-1}(\cdots p_1(\pi) \cdots))$; this boundary path is called the image $\beta(\pi)$ of π.

Our discussion of reflection can be summarized by:

(12) To every auxiliary path π there corresponds a unique image path $\beta(\pi)$, which is a boundary path obtained from π by successive reflections.

Further,

(13) among all the auxiliary paths with the same image path, the number of paths of type $\oplus$ exceeds the number of paths of type $\ominus$ by one.

(13) follows from the fact that the auxiliary paths with the same image path β come in pairs of type ($\oplus$, $\ominus$), as illustrated in figure 1 by paths π_2 and π_3, except for a single "bachelor" path of type $\oplus$ from the origin to one of the three $\oplus$ points immediately next to Γ_l.

The bachelor path of type $\oplus$ is the auxiliary path yielding β after only one reflection; it is uniquely defined for any boundary path β, and is constructed from β as follows. Let v be the last point of contact of β with Γ_l, proceeding along β from the origin in accordance with the directions associated with each of the three step types. (Note that β has at least one point of contact with Γ_l, since β is a boundary path). The bachelor auxiliary path is constructed from β by "reflecting" the portion of β following v. (The word "reflection" is put in quotes because, up to now, reflection has been defined only as an operation on auxiliary paths. But the construction involved here is entirely analogous to the earlier operation.) For example, if v lies in the horizontal side of Γ_l, then "reflection" of the portion of β following v consists of replacing every step of type B by a step of type C, and every step of type C by a step of type B; the procedure is analogous if v lies in one of the other two sides of Γ_l. (Note that v is never a vertex of Γ_l).

The pairing of the other auxiliary paths with image β is accomplished by "reflection" about the last point of contact with the triangular grid lines indicated by the dashed lines in figure 1. (The word "reflection" again is put in quotes, because the usage here does not correspond exactly to the operation

yielding $p(\pi)$ from π). For example, consider an auxiliary path π_2 with image β, and let the last point of contact of π_2 with the triangular grid lines be w; suppose for example that w lies on a grid line oriented in the direction of a step of type B (as illustrated in figure 1). Then, as indicated in figure 1, the mate π_3 of π_2 is obtained from π_2 by replacing every step of type C occurring after w by a step of type A, and every step of type A by a step of type C. The same "reflection" operation, applied to π_3, yields π_2, which establishes the pairing.

That π_2 and its mate π_3 have the same image β is best verified by imagining π_2 and π_3 as undergoing reflection simultaneously.

Except for the single bachelor path, auxiliary paths with the same image thus come in pairs of type ($\oplus$, $\ominus$), except possibly in the case of an auxiliary path, such as that indicated by π_0 in figure 1, whose potential mate π_1 is not one of the auxiliary paths. However, auxiliary paths such as π_0 do not exist, and this is shown as follows.

Suppose there were an auxiliary path, such as π_0, to an endpoint at the outer edge of the hexagonal grid of $\oplus$ points and $\ominus$ points, which entered the triangular cell containing this endpoint from an "exterior" side of the cell. The four equations (10a), (10b), (10c) and (10d) yield $m_C = n - l([n/l])$ for any auxiliary path to any endpoint between the two vertices V_1 and V_2. (Correspondingly $m_B = n - l([n/l])$ and $m_A = n - l([n/l])$ for the other two sets of "outer" endpoints). Hence, if π_0 existed, it would contain $n - l([n/l])$ steps of type C. But then π_1 would contain $n - l([n/l]) - l$ steps of type C, which could not be because $n - l([n/l]) - l$ is negative.

Finally,

(14) Every boundary path is the image of at least one auxiliary path,

because every boundary path is the image at least of its corresponding "bachelor" path.

(12), (13), and (14) imply

$$N_0 = N_\ominus - N_0 . \tag{15}$$

(15) is shown as follows. Let π denote an auxiliary path, let β denote a boundary path, and define the function $f(\pi, \beta)$ as follows.

$$f(\pi, \beta) = 1 \quad \text{if } \beta \text{ is the image of } \pi, \text{ and } \pi \text{ is a path of type } \oplus.$$

$$f(\pi, \beta) = -1 \quad \text{if } \beta \text{ is the image of } \pi, \text{ and } \pi \text{ is a path of type } \ominus.$$

$$f(\pi, \beta) = 0 \quad \text{if } \beta \text{ is not the image of } \pi.$$

Now, for any fixed β,

$$\sum_\pi f(\pi, \beta) = 1$$

by (13) and (14), so that

$$(16) \qquad \sum_{\beta} \left[\sum_{\pi} f(\pi, \beta)\right] = N_0$$

Again, by (12), it is true for every fixed π that

$$\sum_{\beta} f(\pi, \beta) = +1 \text{ for } \pi \text{ of type } \oplus$$

$$= -1 \text{ for } \pi \text{ of type } \ominus,$$

so that

$$(17) \qquad \sum_{\pi} \left[\sum_{\beta} f(\pi, \beta)\right] = N_{\oplus} - N_{\ominus},$$

and (15) follows from (16) and (17).

(11) and (15) yield

$$(18) \qquad N = N_{\oplus} - N_{\ominus}$$

In view of (8), equation (18) represents the solution of the small-sample problem, because the computation of $N_{\oplus}$ and of $N_{\ominus}$ is straightforward. For example, $N_{\oplus}$ is the total number of paths of type $\oplus$, which is easily computed because the number of paths to any particular $\oplus$ point is given by the usual trinomial coefficient, the count being entirely unrestricted. The three arguments of this trinomial coefficient are the numbers of steps of types A, B and C involved in any auxiliary path to this $\oplus$ point; these numbers are of course fixed by the location of the $\oplus$ point, in view of equations (10). There remains only the problem of efficient enumeration of $\oplus$ points and θ points; one such enumeration gives for $\Pr \{D_{3,n} \geqq l/n\}$ the expression

$$(19) \qquad 3 \sum_{i=1}^{[n/l]} \sum_{j \in J(i)} (\pm)(n!)^3/(n - il)!(n + jl)!(n + (i - j)l)!,$$

where the set $J(i)$ consists of the integers $(2 - i, 3 - i, 5 - i, 6 - i, 8 - i, 9 - i, 11 - i, 12 - i, \cdots, 2i)$, and where the $(\pm)$ sign indicates that, for fixed i, successive terms in the finite series indexed by j have alternating signs, beginning with $+$ for $j = 2 - i$, $-$ for $j = 3 - i$, $+$ for $j = 5 - i$, etc.

3. Large-sample distribution. The asymptotic distribution of $D_{3,n}$ is given by the following theorem.

THEOREM. *For $\lambda n^{\frac{1}{2}}$ integral*

$$\lim_{n \to \infty} \Pr \{n^{\frac{1}{2}} D_{3,n} \geqq \lambda\} = 3 \sum_{i=1}^{\infty} \sum_{j \in J(i)} (\pm) e^{-\lambda^2 (i^2 + j^2 - ij)}$$

where the set $J(i)$ and the sign $(\pm)$ are as defined in (19).

PROOF. Put $l = \lambda n^{\frac{1}{2}}$ in (19). Since, for fixed k_1, k_2, k_3 with $k_1 + k_2 + k_3 = 0$,

$$(20) \qquad \lim_{n \to \infty} \frac{(n!)^3}{(n + k_1 n^{\frac{1}{2}})!(n + k_2 n^{\frac{1}{2}})!(n + k_3 n^{\frac{1}{2}})!} = e^{-\frac{1}{2}(k_1^2 + k_2^2 + k_3^2)},$$

suffices to show that,

1)
$$\begin{cases} \text{for } k \text{ large enough,} \\ R(k, n, \lambda) = \left| \sum_{i=k}^{[n^{1/2}/\lambda]} \sum_{j \in J(i)} (\pm)(n!)^3/(n - i\lambda n^{1/2})! \cdot (n + j\lambda n^{1/2})!(n + (i-j)\lambda n^{1/2})! \right| \\ \text{is arbitrarily small, uniformly in } n \text{ for large } n. \end{cases}$$

Rewriting the terms of (21) and putting the absolute value signs inside the st summation,

2)
$$R(k, n, \lambda) \leqq \sum_{i=k}^{[n^{1/2}/\lambda]} ((n!)^3/(n - i\lambda n^{1/2})!(2n + i\lambda n^{1/2})!) \cdot \left(\left| \sum_{j \in J(i)} (\pm) \binom{2n + i\lambda n^{1/2}}{n + j\lambda n^{1/2}} \right| \right).$$

For fixed i, the absolute values of the terms of the alternating series increase onotonically to the maximum

$$\binom{2n + i\lambda n^{1/2}}{n + [i/2]\lambda n^{1/2}},$$

d then decrease monotonically. Hence

$$\left| \sum_{j \in J(i)} (\pm) \binom{2n + i\lambda n^{1/2}}{n + j\lambda n^{1/2}} \right| \leqq 2 \binom{2n + i\lambda n^{1/2}}{n + [i/2]\lambda n^{1/2}},$$

d (22) yields

3)
$$R(k, n, \lambda) \leqq 2 \left[\sum_{i=k}^{n^{1/2}/\lambda} \right] b_i,$$

iere

4)
$$b_i = (n!)^3/(n - i\lambda n^{1/2})!\left(n + \left[\frac{i}{2}\right]\lambda n^{1/2}\right)!\left(n + \left(i - \left[\frac{i}{2}\right]\right)\lambda n^{1/2}\right)!.$$

It is easy to show by direct computation that

1) b_i/b_{i+1} is increasing in i,

2) $b_k/b_{k+1} \geqq \left(1 + \left[\frac{k}{2}\right]\lambda n^{-1/2}\right)^{\lambda n^{1/2}}$, which is uniformly close to

$^{/2]\lambda^2}$ for n large.

ence, by (23), $R(k, n, \lambda)$ is essentially bounded by

5)
$$2b_k/(1 - e^{-[k/2]\lambda^2})$$

for n large. But, by (20) and (24), (25) is approximated by

$$2e^{-(k^2+[k/2]^2-k[k/2])}/(1 - e^{-[k/2]\lambda^2})$$

for n large; this establishes (21).

REFERENCES

[1] S. CHANDRASEKHAR, "Stochastic problems in physics and astronomy", *Selected Papers on Noise and Stochastic Processes* (editor N. Wax), Dover, New York, 1954, pp 4–9.

[2] W. FELLER, *Probability theory and its applications*, Wiley, New York, 1950, p. 304 (problem No. 5)

[3] W FELLER, Review of B. V. Gnedenko, and V S Korolyuk, "On the maximum discrepancy between two empirical distributions", *Mathematical Reviews*, Vol. 13 (1952), pp. 570-571.

[4] M. FISZ, "A limit theorem for empirical distribution functions", *Bulletin de l'Académie Polonaise des Sciences, classe III*, Vol. 5 (1957), pp. 695–698.

[5] I. I. GIHMAN, "On a nonparametric criterion for the homogeneity of k samples", *Teorya Veryoinostei i yeyau primenyenya*, Vol 2 (1957), pp. 380–384

[6] B. V GNEDENKO AND V. S. KOROLYUK, "On the maximum discrepancy between two empirical distributions", *Doklady Akad. Nauk. SSSR (N.S)*, Vol. 80 (1951), pp. 525–528.

[7] J. KIEFER, "Distance tests with good power for the nonparametric k-sample problem" (abstract), *Ann. Math Stat.*, Vol. 26 (1955), p. 775.

[8] V. OZOLS, "Generalization of the theorem of Gnedenko-Korolyuk to three samples in the case of two one-sided boundaries". *Latvijas PSR Zinātnu Akad.* Vēstis 1956, No. 10 (111). pp 141–152 (Listed in *Mathematical Reviews*, Vol. 18 (1957), p 833.)

DISTRIBUTION OF A SERIAL CORRELATION COEFFICIENT NEAR THE ENDS OF THE RANGE[1]

By M. M. Siddiqui

University of North Carolina

1. Introduction and summary. If $y_1, \cdots, y_N$ are observations on a stationary time series at equal intervals of time and it is known that $Ey_t = 0$ for $t = 1, \cdots, N$, the most natural definition of a serial correlation coefficient with lag unity would be

$$r^* = \left(\sum_{i=1}^{N-1} y_i y_{i+1}\right)\left[\left(\sum_{i=1}^{N-1} y_i^2\right)\left(\sum_{i=1}^{N-1} y_{i+1}^2\right)\right]^{-1/2}$$

if the denominator $\neq 0$. This is the ordinary correlation coefficient between $(y_1, \cdots, y_{N-1})$ and $(y_2, \cdots, y_N)$, except that instead of taking deviations from the sample mean, we have taken deviations from the population means. Due to the seemingly unsurmountable mathematical difficulties involved in obtaining the distribution of r^* even on the hypothesis of independence and normality of the observations, several alternative definitions have been proposed as approximations to r^*. However, it is desirable to consider some relevant properties of the distribution of r^*.

In this paper the distribution of r^* near the extremities of its range will be considered. The observations will be assumed to be distributed as independent $N(0, 1)$ variates. There is no loss of generality in assuming the variance to be unity as r^* is independent of the scale parameter. A geometrical approach suggested by Hotelling seemed to be particularly suitable in obtaining the order of contact of the distribution curve at $r^* = \pm 1$. Hotelling [1] shows how to determine the order of contact of frequency curves of some statistics with the variate axis at the ends of the range even though the actual distributions are unknown. It will be shown here that if for a number r_0 in [0, 1] and close to 1, $P(r^* \geqq r_0)$ is expanded in a series of powers of $(1 - r_0)$, the first non-zero coefficient is that of the power $(N - 2)/2$. Upper and lower bounds for the coefficient of this power will be calculated. The lower bound is positive and the upper bound gives an approximation for an upper bound on $P(r^* \geqq r_0)$.

2. Geometrical representation. Let $X_1, \cdots, X_N$ be N independent $N(0, 1)$ variates. Define

$$(2.1) \qquad r^* = \left(\sum X_i X_{i+1}\right)\left[\left(\sum X_i^2\right)\left(\sum X_{i+1}^2\right)\right]^{-1/2}$$

where all the summations are from 1 to $N - 1$ and the denominator $\neq 0$, then r^* is a variate with range [0, 1].

Received August 20, 1957; revised February 7, 1958.

[1] Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States government.

For every set of observations $y_1, \cdots, y_N$ on these variates we take a point S with coordinates $(y_1, \cdots, y_N)$ in an N-dimensional Euclidean space, which may be regarded as a representation of the sample space. Denoting the origin by O, we see that the points S are distributed with spherical symmetry about O. Furthermore, a unique value of r^* corresponds to all the points on a straight line OS, excepting the origin. Let the straight line OS meet the $N - 1$-dimensional unit sphere in Q and Q', where Q is on the same side of the origin with S Denoting by $(x_1, \cdots, x_N)$ the coordinates of Q, we have

$$\sum_{i=1}^{N} x_i^2 = 1, \tag{2.2}$$

which may also be taken as the equation of the unit sphere. The points Q and Q' may be considered to determine a unique value of r^*. Considering only the point Q, it is easily seen that the distribution of Q is uniform over the unit sphere; that is, denoting the total $(N - 1)$-dimensional surface area of (2.2) by S_{N-1}, the probability of Q falling in an area A on the sphere is

$$A/S_{N-1}.$$

For a given r_0 in $[-1, 1]$ there exists a set of points on the unit sphere such that for each point in this set the corresponding value of r^* lies in the interval $[r_0, 1]$, and for no other point. If this set of points covers an area A on the surface of the sphere (2.2), it follows that

$$P(r^* \geqq r_0) = A/S_{N-1}.$$

We observe that $r^* = 1$ if and only if $x_i = \lambda x_{i-1}$, $i = 2, 3, \cdots, N$, $\lambda > 0$ and $x_1 \neq 0$, that is, $x_i = \lambda^{i-1} x_1$, $i = 2, 3, \cdots, N$, $\lambda > 0$ and $x_1 \neq 0$. Since the point $(x_1, \cdots, x_N)$ lies on (2.2), we obtain for the value of x_1, $x_1 = \pm c$ where

$$c = (1 - \lambda^2)^{1/2}(1 - \lambda^{2N})^{-1/2}. \tag{2.3}$$

Denote the variable point $(c, \lambda c, \cdots, \lambda^{N-1}c)$ by P and $(-c, -\lambda c, \cdots, -\lambda^{N-1}c)$ by P'. As λ varies from 0 to ∞, each of P and P' describes a curve for every point of which—excepting the two points of each curve obtained by $\lambda = 0$ and ∞—corresponds the value of $r^* = 1$.

Since both these curves are exactly alike, except for their position in space, we confine our attention to the curve

$$x_1 = c, \qquad x_i = \lambda^{i-1} x_1, \qquad i = 2, \cdots, N, 0 < \lambda < \infty. \tag{2.4}$$

Further, from now on we reserve $(x_1, \cdots, x_N)$ to denote the point on curve (2.4) which corresponds to the parameter λ, and we use $(\epsilon_1, \cdots, \epsilon_N)$ to denote any other point on the unit sphere.

To find the probability of r^* exceeding a given value r_0 which is close to 1, we consider the points within a "tube" of geodesic radius θ on the surface of the sphere (2.2) with its axial curve (2.4).

Let the length of the curve (2.4) measured from $P_0(1, 0, \cdots, 0)$ to

$$P(x_1, \cdots, x_N)$$

be denoted by s, or more explicitly $s(\lambda)$, and an element of curve by ds. Denoting by primes the differential coefficient with respect to s, the direction cosines of the tangent to the curve at P are

$$x_1', x_2', \cdots, x_N',$$

where

(2.5) $$x_i' = [(i-1)\lambda^{i-2}c + \lambda^{i-1}\, dc/d\lambda]\lambda', \qquad i = 1, 2, \cdots, N.$$

We note that

(2.6) $$\sum_{i=1}^{N} x_i'^2 = 1,$$

and since

$$\sum_{i=1}^{N} x_i^2 = 1$$

we have

(2.7) $$\sum_{i=1}^{N} x_i x_i' = 0.$$

Let the coordinate axes be rotated so that the new coordinates are denoted by the elements of a vector α. Let $\alpha = B\epsilon$ where

$$B = \begin{bmatrix} x_1' & x_2' & \cdots & x_N' \\ x_1 & x_2 & \cdots & x_N \\ b_{31} & b_{32} & \cdots & b_{3N} \\ \cdots & \cdots & \cdots & \cdots \\ b_{N1} & b_{N2} & \cdots & b_{NN} \end{bmatrix},$$

and

(2.8) $$BB' = I.$$

Here I denotes the identity matrix, B' denotes the transpose of B, and ϵ and α denote the column vectors $(\epsilon_1, \cdots, \epsilon_N)$ and $(\alpha_1, \cdots, \alpha_N)$ respectively.

The α_1 axis is now parallel to the tangent of the curve at P and the α_2 axis coincides with the line OP.

The $(N-3)$-dimensional sphere given by the set of equations

(2.9) $$\alpha_1 = 0, \qquad \alpha_2 = \cos\theta, \qquad \alpha_i = \beta_i \sin\theta, \qquad i = 3, 4, \cdots, N,$$

with

$$\sum_{i=3}^{N} \beta_i^2 = 1,$$

lies entirely on the $(N-1)$-dimensional unit sphere

$$\sum_{i=1}^{N} \alpha_i^2 = 1 = \sum_{i=1}^{N} \epsilon_i^2. \tag{2.10}$$

The sphere (2.10) is the same as (2.2) with a change of notation. Each point on (2.9) is at a geodesic distance θ from P measured on the sphere (2.10). Further, since (2.9) lies in the plane $\alpha_1 = 0$, this hypersphere is perpendicular to the tangent of curve (2.4) at P.

Changing back to the original coordinates we have $\epsilon = B'\alpha$ or

$$\epsilon_i = x_i'\alpha_1 + x_i\alpha_2 + b_{3i}\alpha_3 + \cdots + b_{Ni}\alpha_N, \qquad i = 1, \cdots, N.$$

Equations (2.9) become

$$\epsilon_j = x_j \cos\theta + \sin\theta \sum_{i=3}^{N} b_{ij}\beta_i, \qquad j = 1, 2, \cdots, N \tag{2.11}$$

with

$$\sum_{i=3}^{N} \beta_i^2 = 1.$$

3. The value of r^* near the curve. Let us calculate the value of r^* corresponding to a point $(\epsilon_1, \cdots, \epsilon_N)$ on the hypersphere (2.11) We have

$$r^* = \left(\sum_{j=1}^{N-1} \epsilon_j \epsilon_{j+1}\right) [(1 - \epsilon_1^2)(1 - \epsilon_N^2)]^{-1/2} \tag{3.1}$$

since

$$\sum_{j=1}^{N-1} \epsilon_j^2 = 1 - \epsilon_N^2 \quad \text{and} \quad \sum_{j=1}^{N-1} \epsilon_{j+1}^2 = 1 - \epsilon_1^2.$$

Inserting the values of ϵ's from (2.11) in terms of x's, using equations (2.4)-(2.8) and neglecting the terms of order $\sin^3\theta$, we have, after some algebraic simplification

$$\begin{aligned} \frac{1 - r^*}{\sin^2\theta} &= 1 + \frac{(1-\lambda^2)(1+\lambda^{2N})}{2\lambda^2(1-\lambda^{2N-2})} + \frac{(1-\lambda^2)(1-\lambda^{2N})}{\lambda^2(1-\lambda^{2N-2})^2} \\ &\cdot \left[\lambda^{N-1} \sum_{k=3}^{N}\sum_{i=3}^{N} b_{i1} b_{kN} \beta_i \beta_k - \frac{\lambda(1-\lambda^{2N-2})}{1-\lambda^2} \sum_{j=1}^{N-1}\sum_{k=3}^{N}\sum_{i=3}^{N} b_{ij} b_{k,j+1} \beta_i \beta_k \right. \\ &\left. - \frac{(1-\lambda^{2N})}{2(1-\lambda^2)} \frac{\left(\sum_{i=3}^{N} b_{i1}\beta_i\right)^2}{\lambda^2} + \lambda^2 \left(\sum_{i=3}^{N} b_{iN}\beta_i\right)^2\right] \end{aligned} \tag{3.2}$$

ere the variable of integration is λ. This can be written as

$$P(r^* \geqq r_0) \doteq \pi^{-1}(1 - r_0)^{(N-2)/2} \int_0^\infty [g(\lambda)]^{-(N-2)/2} h(\lambda)\, d\lambda, \tag{2)}$$

ere

$$g(\lambda) = 1 + \frac{(1-\lambda^2)(1+\lambda^{2N})}{2\lambda^2(1-\lambda^{2N-2})} + \frac{(1-\lambda^{2N-4})(1-\lambda^{2N})}{(N-2)(1-\lambda^{2N-2})^2} \tag{3)}$$

l

$$h(\lambda) = \frac{ds}{d\lambda} = \left[\sum_{i=1}^N \left(\frac{dx_i}{d\lambda}\right)^2\right]^{1/2} = \left[\frac{1}{(1-\lambda^2)^2} - \frac{N^2\lambda^{2N-2}}{(1-\lambda^2)^{2N}}\right]^{1/2}. \tag{4)}$$

note here that E. S. Keeping [4] has studied the integral of $h(\lambda)$ over the ge $[0, \infty]$.

f we change the variable of integration from λ to 1/λ we observe that the egral in (4.2) remains unchanged, hence the integral from 0 to 1 is the same from 1 to ∞. Writing J for the integral in (4 2), we have

$$J = 2\int_0^1 [g(\lambda)]^{-(N-2)/2} h(\lambda)\, d\lambda. \tag{5)}$$

By considering the sign of the differential coefficient of $g(\lambda)$ in the interval 1] we find that $g(\lambda)$ is a monotonically decreasing function of λ, and

$$g(0) = \infty, \qquad g(1) = [N/(N-1)]^2.$$

Write

$$\chi(\lambda) = \frac{1}{g(\lambda)}, \tag{6)}$$

en $\chi(\lambda)$ is a monotonically increasing function of λ in [0, 1] with

$$\chi(0) = 0 \quad \text{and} \quad \chi(1) = (1 - 1/N)^2. \tag{7)}$$

5. Bounds on the integral J. From (4.5) and (4.6) we have

$$J = 2\int_0^1 [\chi(\lambda)]^{(N-2)/2} h(\lambda)\, d\lambda. \tag{1)}$$

w $\chi(\lambda)$ can be written as

$$\chi(\lambda) = 2\lambda^2 \left(\frac{1-\lambda^{2N-2}}{1-\lambda^{2N}}\right)^2 \left[1 + \frac{N\lambda^2(1-\lambda^{2N-4})}{(N-2)(1-\lambda^{2N})}\right]^{-1}. \tag{2)}$$

ake the transformation

$$\lambda = e^{-s/N} \tag{3)}$$

.), then

$$J = \frac{2^{(N-2)/2}}{N} \qquad (5.$$

$$\cdot \int_0^\infty \frac{\left(\cosh \frac{x}{N} - \coth x \sinh \frac{x}{N}\right)^{N-2} \left(\operatorname{cosech}^2 \frac{x}{N} - N^2 \operatorname{cosech}^2 x\right)^{1/2} dx}{\left[1 + \frac{N}{N-2}\left(\cosh \frac{2x}{N} - \coth x \sinh \frac{2x}{N}\right)\right]^{(N-2)/2}}. \qquad (5$$

elementary expansions of hyperbolic functions in power series, for ex-

, e.

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$$

ery x and for $|x| < \pi$,

$$\coth x = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \frac{2x^5}{945} + \cdots,$$

fter some binomial and exponential expansions, we finally obtain

$$/N)]^{(N-2)/2} = e^{-1}\left[1 - \frac{x^2}{6} + \frac{x^4}{40} + \cdots\right]$$

$$+ \frac{3e^{-1}}{2N}\left(1 + \frac{2x^2}{9} - \frac{x^4}{30} + \cdots\right) + O(N^{-2}),$$

this expansion is valid for $|x| < \pi$.

arly

$$(e^{-x/N})\, d(e^{-x/N}) = -\frac{1}{2\sqrt{3}}\left[1 - \frac{x^2}{10} + \frac{135}{12600}x^4 + \cdots + O(N^{-2})\right] dx.$$

split the range of integration in (5.4) into the ranges [0, 1] and [1, ∞]. ting the integral from 0 to 1 by J_1, we have, omitting the terms $O(N^{-1})$,

$$J_1 = 3^{-1/2}e^{-1} \int_0^1 \left(1 - \frac{x^2}{6} + \frac{x^4}{40} + \cdots\right)$$

$$\cdot\left(1 - \frac{x^2}{10} + \frac{137}{12600}x^4 + \cdots\right) dx = 3^{-1/2}e^{-1}(.9216) = 0.196.$$

e

$$J = 0.196 + 2\int_0^{e^{-1/N}} [\chi(\lambda)]^{(N-2)/2} h(\lambda)\, d\lambda.$$

note by J_2 the second term on the right hand side of (5.6) and substitute $y^{1/2}$ so that

$$J_2 = 2^{(N-2)/2} \int_0^{e^{-2/N}} \frac{y^{(N-3)/2}}{(1-y)} \left(\frac{1-y^{N-1}}{1-y^N}\right)^{N-2} \tag{5.7}$$
$$\cdot \left[1 + \frac{Ny(1-y^{N-2})}{(N-2)(1-y^N)}\right]^{-(N-2)/2} \left[1 - \frac{N^2 y^{N-1}(1-y)^2}{(1-y^N)^2}\right]^{1/2} dy.$$

It can easily be shown that

$$\frac{e^{-1/2}}{(1+y)^{(N-2)/2}} < \left[1 + \frac{Ny(1-y^{N-2})}{(N-2)(1-y^N)}\right]^{-(N-2)/2} < \frac{1}{(1+y)^{(N-2)/2}}. \tag{5.8}$$

The other factors in the integrand can be expanded by the binomial theorem, e.g.

$$\left[1 - \frac{N^2 y^{N-1}(1-y)^2}{(1-y^N)^2}\right]^{1/2} = 1 - \frac{N^2 y^{N-1}(1-y)^2}{2(1-y^N)^2} - \frac{N^4 y^{2N-2}(1-y)^4}{8(1-y^N)^4} - \cdots$$

We then have

$$e^{-1/2} Q < J_2 < Q, \tag{5.9}$$

where

$$Q = 2^{(N-2)/2} \int_0^{e^{-2/N}} \frac{y^{(N-3)/2}\, dy}{(1+y)^{(N-2)/2}} \tag{5.10}$$
$$\cdot \left[\frac{1}{1-y} - (N-2)y^{N-1} - \frac{N^2}{2} \frac{y^{N-1}(1-y)}{1-y^N} + \cdots\right].$$

We observe that we have to evaluate integrals of type

$$M(p, q, e^{-2/N}) = \int_0^{e^{-2/N}} y^p (1+y)^{-q}\, dy \tag{5.11}$$

and

$$L(p, q, e^{-2/N}) = \int_0^{e^{-2/N}} y^p (1+y)^{-q} (1-y)^{-1}\, dy, \tag{5.12}$$

where $q = (N-2)/2$ and $p = sq + b$, $s > 0$.

Substituting $y = e^{-2/N} z$ and expanding $(1 + ze^{-2/N})^{-q}$ in powers of $(1-z)$ and integrating term by term, we obtain

$$M(p, q, e^{-2/N}) = \frac{e^{-2(p+1)/N}}{(p+1)(1+e^{-2/N})^q} F\left(1, q, p+2, \frac{1}{1+e^{2/N}}\right)$$

and

$$L(p, q, e^{-2/N}) = \frac{e^{-2(p+1)/N}}{(1+e^{-2/N})^q} \sum_{k=0}^{\infty} F\left(1, q, p+k+2, \frac{1}{1+e^{2/N}}\right).$$

If $s > 1$, $b > 0$ and $x > 0$

$$F(1, q, sq + b, x) = 1 + \frac{q}{sq + b} x + \frac{q(q + 1)}{(sq + b)(sq + b + 1)} x^2 + \cdots$$
$$> 1 + \frac{q}{sq + b} x + \frac{q^2}{(sq + b)^2} x^2 + \cdots$$
$$= \left(1 - \frac{qx}{sq + b}\right)^{-1}$$

and

$$F(1, q. sq + b, x) < 1 + \frac{q}{sq + b} x$$
$$+ \frac{q(q + 1)}{(sq + b)(sq + b + 1)} x^2[1 + x + x^2 + \cdots].$$

Since $q = O(N)$, omitting the terms of $O(N^{-1})$ we have

$$\frac{2s}{2s - 1} < F\left(1, q, sq + b, \frac{1}{1 + e^{2/N}}\right) < \frac{1 + s + 2s^2}{2s^2}.$$

A systematic calculation then shows that

$$\frac{.542}{2^{(N-2)/2}} [1 + O(N^{-1})] < L(p, q, e^{-2/N}) < \frac{.629}{2^{(N-2)/2}} [1 + O(N^{-1})].$$

Denoting the integrals of successive terms in (5.10) by Q_1, Q_2, etc., as they occur in order and neglecting the sign, we see that

$$Q_1 = 2^{(N-2)/2} L\left(\frac{N - 3}{2}, \frac{N - 2}{2}, e^{-2/N}\right).$$

Hence

$$0.542 < Q_1 < 0.629.$$

Similar calculations on the following terms show that

$$Q < .629 - .065 - .101 + .029 - .005 = .487$$

and

$$Q > .542 - .066 - .103 + .028 - .006 = .395.$$

The terms diminish very rapidly and the later terms do not affect the second decimal place. Thus from (5.9)

$$.239 < J_2 < .487,$$

and since

$$J = J_1 + J_2 = .196 + J_2$$

therefore

$$.435 < J < .683. \tag{5.13}$$

These calculations are valid to two decimal places and $O(N^{-1})$. Finally, the first term, P_0, in the expansion of $P(r^* \geqq r_0)$ in powers of $(1 - r_0)$ is

$$P_0 = J/\pi(1 - r_0)^{(N-2)/2}.$$

It is easy to see that the first term in the expansion of $P(r^* \leqq -r_0)$ is the same as P_0.

If the population mean is known to be zero, the frequency function of the ordinary correlation coefficient, r, for a sample of size N is given by

$$f(r) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{N-2}{2}\right)}(1 - r^2)^{(N-4)/2}.$$

Therefore the first term in the expansion of $P(r \geqq r_0)$ in powers of $(1 - r_0)$ is approximately

$$P \doteq 2^{(N-3)/2}\pi^{-\frac{1}{2}}(N - 2)^{-\frac{1}{2}}(1 - r_0)^{(N-2)/2}.$$

Hence

$$P_0/P \doteq 2^{-(N-3)/2}(N - 2)^{\frac{1}{2}}\pi^{-\frac{1}{2}}J,$$

which tends to zero as N tends to infinity.

6. Acknowledgement. The author wishes to express his indebtedness to Professor Harold Hotelling for suggesting this problem and for many helpful suggestions during the preparation of this paper.

REFERENCES

[1] HAROLD HOTELLING, "An application of analysis situs to statistics," *Bull. Amer. Math. Soc.*, Vol. 33(1927), pp. 467–476.
[2] HAROLD HOTELLING, "Tubes and spheres in n-spaces and a class of statistical problems," *Amer. J. Math.*, Vol. 61 (1939), pp. 440–460.
[3] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1951.
[4] E. S. KEEPING, "A significance test for exponential regression," *Ann. Math Stat*, Vol. 22(1950), pp. 180–198.

DISTRIBUTIONS OF THE MEMBERS OF AN ORDERED SAMPLE

By Charles E. Clark and G. Trevor Williams

Booz, Allen and Hamilton, and The Johns Hopkins University

1. Introduction. Let the members of a random sample from a distribution $F(x)$ with probability density $F'(x) = f(x)$ be in order of magnitude $x_1, \cdots, x_m, \cdots, x_n, \cdots, x_N$, with $x_i \leqq x_{i+1}$, $i = 1, \cdots, N - 1$, and $m < n$. We shall compute the moments of the distribution of x_m and of the joint distribution of x_m and x_n.

The results are derived under the assumption that $F^{-1}(x)$, the inverse of $F(x)$, is a polynomial. Then we discuss the applicability of the results to any distribution for which $F^{-1}(x)$ is differentiable at $m/(N + 1)$ and $n/(N + 1)$. In this general case no restriction on $F(x)$ is imposed other than the differentiability; in particular, the interval on which $0 < F(x) < 1$ can be finite, semi-finite, or infinite.

2. Present status of the problem. This problem is handled through analyses of several specific distributions in reference [1] listed at the end of this paper. It is suggested that any one of the Pearson type frequency curves can be adequately approximated by one of the density functions handled in that paper. Although a general method is employed, there is no general development or general results; each distribution requires special, extensive computations. In contrast to these earlier results, the present paper contains a general development with results that are easily specialized to particular distributions.

Following [1] there have been discussions of asymptotic distributions. It is known that if m and N increase with m/N approaching a limit different from zero and one, under quite general conditions the distribution of x_m is asymptotically normal; see [2] or [3]. Also it was pointed out in [4] that with some restrictions on the distribution function the limiting distribution of x_m as N increases, but m is fixed, has the probability density

$$m^m \exp\,[my - \exp\,(-y)]/(m - 1)!$$

where y is a normalization of x_m; see [5]. However, it is suggested in [6] that in the case of the normal distribution if $m = 1$, one should have a sample of size 10^{12}, and Mr. Kendall concludes in [5], p. 221, that "For practical purposes, therefore, there is still no adequate general approximate form for the distribution of mth values." However, a contribution to the asymptotic case of this problem is made in [6]. In contrast to these asymptotic results, the present paper is concerned with the exact sampling distributions for any sample size. In the case of large samples, known approximations concerning moments are equivalent to the leading terms of some of the expansions of this paper.

Received August 20, 1957; revised December 26, 1957.

3. The moments of the distribution of x_m. The probability density function of x_m is

$$[B(m, N - m + 1)]^{-1}[F(x)]^{m-1}[1 - F(x)]^{N-m}f(x) \tag{1}$$

where the coefficient is the reciprocal of the beta function.

We shall use the random variable $t = F(x_m)$ whose probability density function is $[B(m, N - m + 1)]^{-1}t^{m-1}(1 - t)^{N-m}$. We denote the central moments of this distribution by ν_i, $i = 0, 1, 2, \cdots$. Using p to denote the mean, we compute that $p = m/(N + 1)$.

At first we shall assume that the inverse of $F(x)$ is

$$F^{-1}(x) = \sum_{i=0}^{r} a_i(x - p)^i. \tag{2}$$

Later we shall remove the restriction that $F^{-1}(x)$ is a polynomial.

The kth raw moment of the distribution of x_m immediately reduces to

$$\mu_k' = [B(m, N - m + 1)]^{-1} \int_0^1 [F^{-1}(t)]^k t^{m-1}(1 - t)^{N-m}\, dt, \qquad k = 0, 1, \cdots.$$

For each k we can write as a finite sum

$$[F^{-1}(t)]^k = \sum b_i(t - p)^i$$

where the coefficients b_i are functions of a_i and k. In this notation we have

$$\mu_k' = \sum b_i \nu_i . \tag{3}$$

We calculate that

$$\nu_i = \sum_{j=0}^{i} \binom{i}{j} \frac{N!}{(m - 1)!(N - m)!} (-p)^j \int_0^1 t^{i-j+m-1}(1 - t)^{N-m}\, dt$$

$$= p^i \sum_{j=0}^{i-2} (-1)^j \binom{i}{j} \frac{(m + 1)\cdots(m + i - j - 1)}{p^{i-j-1}(N + 2)\cdots(N + i - j)} + (-1)^{i-1} i p^i + (-p)^i.$$

This expression will be reduced to a more convenient form. We use the identity

$$\frac{m + A}{p(N + A + 1)} = 1 + \frac{Aqp^{-1}}{N + A + 1}, \qquad q = 1 - p = \frac{N - m + 1}{N + 1}$$

and reduce ν_i to

$$\nu_i = p^i \sum_{j=0}^{i} (-1)^j \binom{i}{j} \prod_{A=1}^{i-j-1} \left(1 + \frac{Aqp^{-1}}{N + A + 1}\right), \qquad i = 0, 1, 2, \cdots.$$

In this formula

$$\prod_{A=1}^{-1} \left(1 + \frac{Aqp^{-1}}{N + A + 1}\right) = \prod_{A=1}^{0} \left(1 + \frac{Aqp^{-1}}{N + A + 1}\right) = 1.$$

From this result we get $\nu_0 = 1$, $\nu_1 = 0$, and

$$\nu_2 = \frac{pq}{N+2},$$

$$\nu_3 = \frac{2pq(q-p)}{(N+2)(N+3)},$$

$$\nu_4 = \frac{3p^2q^2N + 3pq(2-5pq)}{(N+2)(N+3)(N+4)},$$

$$\nu_5 = \frac{20p^2q^2(q-p)N + 4pq(q-p)(6+5pq)}{(N+2)(N+3)(N+4)(N+5)},$$

$$\nu_6 = \frac{15p^3q^3N^2 + 10p^2q^2(13-40pq)N + 5pq(24-94pq+37p^2q^2)}{(N+2)(N+3)(N+4)(N+5)(N+6)}.$$

We shall use the notation $x_p = F^{-1}(p)$, and $f^{(i)} = f^{(i)}(x_p)$. We can express the a_i in (2) in terms of the derivatives of $F(x)$ at x_p by means of the relations between the derivatives of a function and its inverse. From the a_i we calculate the b_i, and with the use of (3) we get the raw moments. These include

$$\mu_1' = x_p - \frac{f'}{2f^3} \cdot \frac{pq}{N+2} + \frac{3f'^2 - ff''}{6f^5} \cdot \frac{2pq(q-p)}{(N+2)(N+3)} + \frac{10ff'f'' - f^2f''' - 15f'^3}{24f^7} \cdot \frac{3p^2q^2N + 3pq(2-5pq)}{(N+2)(N+3)(N+4)} + \cdots.$$

Here as elsewhere derivatives are denoted by primes and powers by arabic numerical exponents. Finally the central moments μ_k are obtained, such as the following.

$$\mu_2 = \frac{1}{f^2} \cdot \frac{pq}{N+2} - \frac{f'}{f^4} \cdot \frac{2pq(q-p)}{(N+2)(N+3)} + \left[\frac{5f'^2}{4f^6} - \frac{f''}{3f^5}\right] \frac{3p^2q^2N + 3pq(2-5pq)}{(N+2)(N+3)(N+4)} - \frac{f'^2}{4f^6} \cdot \frac{p^2q^2}{(N+2)^2} + \cdots,$$

$$\mu_3 = \frac{1}{f^3} \cdot \frac{2pq(q-p)}{(N+2)(N+3)} - \frac{3f'}{2f^5} \cdot \frac{3p^2q^2N + 3pq(2-5pq)}{(N+2)(N+3)(N+4)} + \frac{3f'}{2f^5} \cdot \frac{p^2q^2}{(N+2)^2} + \cdots,$$

$$\mu_4 = \frac{1}{f^4} \cdot \frac{3p^2q^2N + 3pq(2-5pq)}{(N+2)(N+3)(N+4)} + \cdots.$$

From these results we check the well known fact that if N increases with m/N fixed, the asymptotic distribution of x_m has the mean and variance x_p and pq/f^2N respectively (see [3]). Furthermore the known result that for large N the distribution is approximately normal is suggested by the following which are obtained from the leading terms of the above expressions.

$$\frac{\mu_3}{\mu_2^{3/2}} = N^{-1/2}\left[\frac{2(q-p)}{\sqrt{pq}} - \frac{3f'\sqrt{pq}}{f^2}\right] + \cdots,$$

$$\frac{\mu_4}{\mu_2^2} = 3\left[1 - \frac{5N+12}{(N+3)(N+4)}\right] + \cdots.$$

We next discuss the applicability of the results to distributions for which $F^{-1}(x)$ is not a polynomial. We note that the factor

$$[F(x)]^{m-1}[1 - F(x)]^{N-m}$$

in (1) assumes its maximum value at $(m - 1)/(N - 1)$. Hence (1) indicates that the probability density of x_m is practically zero except in a small neighborhood of $F^{-1}[(m - 1)/(N - 1)]$.[1] Hence the moments of the distribution of x_m can be determined with great accuracy from a knowledge of $F(x)$ in a small neighborhood of $F^{-1}[(m - 1)/(N - 1)]$. But this knowledge of $F(x)$ is given by a few derivatives of $F(x)$ at x_p because x_p is near

$$F^{-1}[(m - 1)/(N - 1)].$$

In other words, the first few terms of the Taylor expansion of $F^{-1}(x)$ at x_p should be enough to permit an accurate determination of the moments. Hence the above derivation holds with very little error if (2) is understood to be a few terms of the Taylor expansion.

4. The median. The results simplify in the case $N = 2m + 1$. We can compute that

$$\int_0^1 (t - 1/2)^j t^m (1 - t)^m \, dt,$$

which is clearly zero when j is odd, reduces when j is even to

$$\frac{m!}{2^{m+j}(j+1)(j+3)\cdots(j+2m+1)},$$

the reduction is achieved by the substitution of $t = \sin^2 \theta$ and use of a known integral (see [7]). This reduces, after multiplication by $B[(m + 1, m + 1)]^{-1}$, to

$$\nu_{2i} = \frac{1\cdot3\cdot5\cdots(2i-1)}{4^i(2m+3)(2m+5)\cdots(2m+2i+1)}, \qquad i = 1, 2, \cdots.$$

5. The efficiency of the median. As a numerical illustration we shall compute the efficiency of the median as an estimator of the mean of a normal distribution. We consider $\varphi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ and $\varphi = \varphi(0)$. The derivatives of $F^{-1}(x)$ at $x = 0$ are calculated from those of $\varphi(x)$. Using (3) with $k = 1, 2$ and the formulas of section 4, we obtain the variance of the median of a sample of size

[1] This statement is true even when $m - 1$ or $N - m$ is small. If, for example, $m - 1$ is small, $F(x) < (m - 1)/(N - 1)$ for $x < F^{-1}[(m - 1)/(N - 1)]$, and $[1 - F(x)]^{N-m}$ is clearly small if x is at least a little greater than $F^{-1}[(m - 1)/(N - 1)]$.

$N = 2n + 1$ in the form

$$\mu_2 = \frac{1}{4\varphi^2(2n+3)}\left\{1 + \frac{1}{4\varphi^2(2n+5)} + \frac{13}{96\varphi^4(2n+5)(2n+7)} + \frac{287}{2688\varphi^6(2n+5)(2n+7)(2n+9)} + \cdots\right\}.$$

Since the sample mean is efficient, and since the variance of the sample mean is $1/(2n + 1)$, if $E(2n + 1)$ is the efficiency of the median,

$$E(2n+1) = [(2n+1)\mu_2]^{-1}.$$

Evaluating φ we obtain

$$\frac{1}{E(2n+1)} = \frac{1.5707963(2n+1)}{2n+3} \cdot \left\{1 + \frac{1.5707963}{2n+5} + \frac{5.3460357}{(2n+5)(2n+7)} + \frac{26.484528}{(2n+5)(2n+7)(2n+9)} + \cdots\right\}.$$

A tabulation of this four term approximation appears in Table I.

The series for the reciprocal of the efficiency converges slowly for small

$$2n + 1.$$

In cases $n = 1, 2, 3$, the fourth term contributes 2.8%, 1.6%, 1.0%, respectively, of the tabulated value. To check the accuracy of the approximation we have calculated accurately (as described below) the reciprocal of the efficiency in cases $n = 1, 2, 3$. The values correct to three decimal places are given in the table. The relative errors are 5.6%, 2.2%, 1.1%, respectively.

TABLE I

Efficiency of the Median, Normal Distribution

$N = 2n + 1$	$[E(2n+1)]^{-1}$, four term approximation	$[E(2n+1)]^{-1}$, exact	$E(2n+1)$
∞	1.571	1.571	.637
201	1.567		.638
101	1.564		.639
51	1.557		.642
31	1.549		.646
21	1.538		.650
11	1.503		.665*
9	1.486		.673*
7	1.457	1.473	.679
5	1.402	1.434	.697
3	1.270	1.346	.743

The third decimal places in E(11) and E(9) are in doubt.

The correct values of the reciprocal of the efficiency are obtained as follows. If $n = 1$, the reciprocal of the efficiency is, except for the factor

$$(2n + 1)/B(2, 2) = 18,$$

with $F'(x) = \varphi(x)$,

$$\int_{-\infty}^{\infty} x^2 F(1 - F)\varphi\, dx = \int_{-\infty}^{\infty} F\, d(-x\varphi + F) - \int_{-\infty}^{\infty} F^2\, d(-x\varphi + F)$$

$$= 1 - \int_{-\infty}^{\infty} (-x\varphi + F)\varphi\, dx - 1 + \int_{-\infty}^{\infty} (-x\varphi + F)2F\varphi\, dx$$

$$= -\left[\frac{\varphi^2}{2} + \frac{F^2}{2}\right]_{-\infty}^{\infty} + \left[\frac{2F^3}{3}\right]_{-\infty}^{\infty} + 2\int_{-\infty}^{\infty} F\, d\left(\frac{\varphi^2}{2}\right)$$

$$= -1/2 + 2/3 - 2\int_{-\infty}^{\infty} \frac{\varphi^2}{2}\varphi\, dx$$

$$= 1/6 - (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-3x^2/2}\, dx$$

$$= 1/6 - \frac{1}{2\pi\sqrt{3}}.$$

Multiplying this last number by $3/B(2, 2)$ we get

$$\frac{1}{E(3)} = 3 - \frac{3\sqrt{3}}{\pi}$$

$$= 1.346$$

as given above.

For $n = 2, 3$ the reciprocals of the efficiencies were calculated by numerical evaluation of

$$\frac{(2n + 1)}{B(n + 1, n + 1)} \int_{-\infty}^{\infty} x^2 F^n (1 - F)^n\, dx.$$

6. The moments of the joint distribution of x_m and x_n, $m < n$. We consider next the joint distribution of x_m and x_n, $m < n$. The probability density is

$$\frac{N!}{(m - 1)!(n - m - 1)!(N - n)!} \cdot [F(x_m)]^{m-1}[F(x_n) - F(x_m)]^{n-m-1}[1 - F(x_n)]^{N-n} f(x_m) f(x_n).$$

The probability density of $t = F(x_m)$ and $u = F(x_n)$ is

$$\frac{N!}{(m - 1)!(n - m - 1)!(N - n)!} t^{m-1}(u - t)^{n-m-1}(1 - u)^{N-n}.$$

The expected values of t and u are $p_m = m/(N + 1)$ and $p_n = n/(N + 1)$ respectively. If $\nu_{\alpha\beta}$ is the expected value of $(t - p_m)^\alpha(u - p_n)^\beta$, we calculate that

$$\nu_{20} = \frac{p_m q_m}{N + 2},$$

$$\nu_{11} = \frac{p_m q_n}{N + 2},$$

$$\nu_{02} = \frac{p_n q_n}{N + 2},$$

$$\nu_{30} = \frac{2p_m q_m(q_m - p_m)}{(N + 2)(N + 3)},$$

$$\nu_{21} = \frac{2p_m q_n(q_m - p_m)}{(N + 2)(N + 3)},$$

$$\nu_{12} = \frac{2p_m q_n (q_n - p_n)}{(N + 2)(N + 3)},$$

$$\nu_{03} = \frac{2p_n q_n (q_n - p_n)}{(N + 2)(N + 3)},$$

$$\nu_{40} = \frac{3p_m^2 q_m^2 N + 3p_m q_m (2 - 5p_m q_m)}{(N + 2)(N + 3)(N + 4)},$$

$$\nu_{31} = \frac{3p_m^2 q_m q_n N + 3p_m q_n (2 - 5p_m q_m)}{(N + 2)(N + 3)(N + 4)},$$

$$\nu_{22} = \frac{p_m q_n[1 - (p_m + q_n) + 3p_m q_n]N + p_m q_n[1 + 5(p_m + q_n) - 15p_m q_n]}{(N + 2)(N + 3)(N + 4)},$$

$$\nu_{13} = \frac{3p_m p_n q_n^2 N + 3p_m q_n (2 - 5p_n q_n)}{(N + 2)(N + 3)(N + 4)},$$

$$\nu_{04} = \frac{3p_n^2 q_n^2 N + 3p_n q_n (2 - 5p_n q_n)}{(N + 2)(N + 3)(N + 4)}.$$

If $\mu'_{\alpha\beta}$ is the expected value of $x_m^\alpha x_n^\beta$,

$$\mu'_{\alpha\beta} = \frac{N!}{(m - 1)!(n - m - 1)!(N - n)!} \int_0^1 du \cdot \int_0^u [F^{-1}(t)]^\alpha [F^{-1}(u)]^\beta \, t^{m-1}(u - t)^{n-m-1}(1 - u)^{N-n} \, dt.$$

Let the Taylor expansion

$$[F^{-1}(t)]^\alpha [F^{-1}(u)]^\beta = a_{00} + a_{10}(t - p_m) + a_{01}(u - p_n) + a_{20}(t - p_m)^2 + a_{11}(t - p_m)(u - p_n) + a_{02}(u - p_n)^2 + \cdots$$

be finite. Then

$$\mu'_{\alpha\beta} = a_{00} + a_{20}\nu_{20} + a_{11}\nu_{11} + a_{02}\nu_{02} + a_{30}\nu_{30} + \cdots .$$

The coefficients a_{ij} are expressed in terms of the derivatives of $F(x)$ at $F^{-1}(p_m)$ and $F^{-1}(p_n)$.

As in the 1-dimensional case, if the Taylor expansion does not terminate, these results are approximations.

As an illustration of the results obtained in this manner, the covariance of x_m and x_n reduces to

$$\begin{aligned} V(x_m, x_n) = {} & \frac{1}{f_m f_n} \cdot \frac{p_m q_n}{N+2} - \frac{f'_m}{2f_m^3 f_n} \cdot \frac{2p_m q_n (q_m - p_m)}{(N+2)(N+3)} \\ & - \frac{f'_n}{2f_n^3 f_m} \cdot \frac{2p_m q_n (q_n - p_n)}{(N+2)(N+3)} \\ & + \frac{3f'^2_m - f_m f''_m}{6f_m^5 f_n} \cdot \frac{3p_m^2 q_m q_n N + 3p_m q_n (2 - 5p_m q_m)}{(N+2)(N+3)(N+4)} \\ & + \frac{3f'^2_n - f_n f''_n}{6f_n^5 f_m} \cdot \frac{3p_m p_n q_n^2 N + 3p_m q_n (2 - 5p_n q_n)}{(N+2)(N+3)(N+4)} \\ & + \frac{f'_m f'_n}{4f_m^3 f_n^3} \cdot \frac{p_m q_n[1 - (p_m + q_n) + 3p_m q_n]N + p_m q_n[1 + 5(p_m + q_n) - 15p_m q_n]}{(N+2)(N+3)(N+4)} \\ & - A_m A_n + \cdots \end{aligned}$$

where

$$f_m^{(i)} = f^{(i)}[F^{-1}(p_m)], \quad i = 0, 1, \cdots,$$

$$A = -\frac{f'}{2f^3} \cdot \frac{pq}{N+2} + \frac{3f'^2 - f''}{6f^5} \cdot \frac{2pq(q-p)}{(N+2)(N+3)} + \frac{10ff'f'' - f^2 f''' - 15f'^3}{24f^7} \cdot \frac{3p^2 q^2 N + 3pq(2 - 5pq)}{(N+2)(N+3)(N+4)},$$

A_m is obtained from A by affixing the subscript m to every f, p, and q, and A_n is obtained similarly.

Using μ_2 as calculated above, we obtain from the last result the first two terms of the coefficient of linear correlation in the form

$$r(x_m, x_n) = \left(\frac{p_m q_n}{q_m p_n}\right)^{1/2} \left\{1 - \frac{A}{N+2}\right\}$$

in which

$$A = \frac{f'^2_m}{4f_m^4} p_m q_m - \frac{f'_m f'_n}{2f_m^2 f_n^2} p_m q_n + \frac{f'^2_n}{4f_n^4} p_n q_n .$$

The following special cases are easily obtained. If $f(x) = \exp(-x)$,

$$A = \tfrac{1}{4}[p_m q_m \exp(2x_m) - 2p_m q_n \exp(x_m + x_n) + p_n q_n \exp(2x_n)].$$

If $f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$,

$$A = \frac{x_m^2}{4f_m^2} p_m q_m - \frac{x_m x_n}{2f_m f_n} p_m q_n + \frac{x_n^2}{4f_n^2} p_n q_n .$$

If $f(x) = \exp(-x)x^{r-1}/\Gamma(r)$,

$$\begin{aligned} A = \tfrac{1}{4}[\Gamma(r)]^2[(r - 1 - x_m)^2 x_m^{-2r} \exp(2x_m) p_m q_m \\ - 2(r - 1 - x_m)(r - 1 - x_n)(x_m x_n)^{-r} \exp(x_m + x_n) p_m q_n \\ + (r - 1 - x_n)^2 x_n^{-2r} \exp(2x_n) p_n q_n . \end{aligned}$$

REFERENCES

[1] K. Pearson, and M. V. Pearson, "On the mean character and variance of a ranked individual and on the mean and variance of the intervals between ranked individuals," Part 1, *Biometrika*, 23 (1931), pp. 364–397, and part 2, *Biometrika*, 24 (1932), pp. 203–279.

[2] N. Smirnoff, "Ueber die Verteilung des allgemeinen Gliedes in der Variationsreihe," *Metron*, 12 (1931), pp. 127–138.

[3] H. Cramér, "Mathematical methods of statistics," Princeton, 1946, pp. 367–378.

[4] E. J. Gumbel, "Les valeurs extremes des distributions statistiques," *Annales de l'Institut Henri Poincare*, 5, (1934), p. 115.

[5] M. G. Kendall, "The advanced theory of statistics," Volume 1, London, 1943, pp. 218–224.

[6] R. A. Fisher, and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest number of a sample," *Proc. Camb. Phil. Soc.*, 24 (1928), p. 180.

[7] E. T. Whittaker and G. N. Watson, "Modern analysis," Cambridge University Press, 1945, p. 256.

POWER FUNCTION CHARTS FOR SPECIFYING NUMBERS OF OBSERVATIONS IN ANALYSES OF VARIANCE OF FIXED EFFECTS

By Leonard S. Feldt and Moharram W. Mahmoud

State University of Iowa and Egyptian Ministry of Education

1. Summary. The charts presented in this paper are designed to facilitate the estimation of the number of observations per treatment required for analysis of variance tests of specified power. They are intended for use by experimenters dealing with fixed treatments effects. With these charts the experimenter may answer the following question: How many observations must I use per treatment to obtain a power of P_1 against alternative hypothesis H_1? Charts are presented for use with tests of treatments effects involving two to five levels of the treatment variable. The charts are strictly valid only for the completely randomized design; however they may be applied with relatively little error to tests of treatments effects in the randomized block and factorial designs, the latter employing a within-cells estimate of error variance.

2. Nature of the charts. Charts are presented for α equal to .05 and .01 and k, the number of levels of the treatment variable, equal to 2, 3, 4 and 5. The charts are entered with the parameter λ, which is defined as follows:

$$\lambda = \sqrt{\frac{\sum_{j}^{k} (\mu_j - \mu)^2}{k\sigma^2}}$$

where μ_j is the mean of treatment population j, μ the mean of the combined treatment populations, σ^2 the population error variance, and k the number of treatments. The value of n, the number of observations which will be required per treatment for a test of specified power, is read directly from the ordinate of the appropriate chart. It is assumed that the same number of observations will be used in every treatment. The relation of λ to ϕ, the parameter customarily employed in the definition of the power function of the F-test, is simply

$$\lambda = \phi\sqrt{n}.$$

3. Historical development. The first extensive tables of the power function of analysis of variance tests were published by Tang [5]. The tables given by Tang are designed in such a way that for fixed values of α, ϕ, f_1 (degrees of freedom for treatments) and f_2 (degrees of freedom for error) the probability of a Type II error may be determined. The interval of tabulation of Tang's tables is .50, however, which is not sufficiently fine for accurate interpolation.

Following Tang's procedure, Lehmer [3] tabulated the values of ϕ for α =

Received December 17, 1957; revised January 31, 1958.

.05 and .01, $P = .7$ and .8, over a wide range of f_1 and f_2. These tables are quite complete within the power range considered, however they can not be conveniently used in the planning of experiments. From the tables the experimenter can tell only that a projected test will have a power less than .7, between .7 and .8, or greater than .8 against a specified alternative.

Pearson and Hartley [4] presented families of power curves for various combinations of α, f_1, and f_2 which make possible a direct estimate of the power of analysis of variance tests. These curves, like the tables of Tang, are entered with the parameter ϕ. For any given experimental setup, the power of the test may be read directly from the ordinate of the curve. These charts are well suited to the evaluation of the power of any given test. They can not be easily employed, however, to indicate the value of n which should be adopted in order to secure a specified power. For this purpose, the experimenter must adopt the relatively inefficient approach of making repeated approximations until the value of n has been estimated with sufficient accuracy.

Fox [2] contributed charts which facilitate the determination of sample size. These charts were constructed from the tables of Tang and Lehmer and are essentially graphs of constant ϕ for varying values of f_1 and f_2. By a method of successive approximations, the value of n may be determined for a fixed value of α and a fixed value of P against a specified alternative. These charts are somewhat more convenient than the curves of Pearson and Hartley for this purpose, but they are somewhat laborious to use because of the iterative nature of the method of approximating n. Also, the charts do not extend below $f_1 = 3$. For experimenters dealing with fixed treatments effects, this limitation considerably restricts their usefulness.

Duncan [1] published a special condensation of the Pearson and Hartley charts. He plotted on a single set of axes the values of ϕ corresponding to $P = .50$ and .90 for various values of f_1 and f_2. Separate charts are presented for $\alpha = .05$ and .01. Having f_1 and f_2 on the same chart facilitates computations which involve both of these elements. For use in planning experiments, however, these charts are subject to the same weaknesses as those of Pearson and Hartley

Though several types of charts and tables of the power function of F-tests have been published, none permits a direct, non-iterative approximation for the number of observations required for a test of specified power. The charts presented in this paper make possible such an approximation for experiments which include 2 to 5 levels of the treatment variable.

4. Construction of the charts. Each chart presents, for $\alpha = .05$ and .01, a family of five curves which correspond to the following values of P: .5, .7, .8, .9 and .95. The number of observations per treatment (n) is plotted on the ordinate, the value of λ is plotted on the abscissa.

The numerical calculations for the coordinates of the points on the curves .7 and .8 were carried out from the tables of Lehmer; the calculations for the remaining curves were based on data read from the charts of Pearson and Hartley. The three basic steps in the calculations were as follows:

(1) Determine (from table or chart) pairs of values for ϕ and f_2 for specified value of P, f_1 and α.
(2) Solve f_2 for n from the relationship $n = 1 + f_2/k$, where k is the number of treatments and n the number of observations per treatment.
(3) Divide ϕ by $\sqrt{n}$ to obtain λ.

The pairs of coordinates, n and λ, were then plotted and smooth curves fitted through these points.

5. Example. An experimenter wishes to investigate the legibility of two common styles of handwriting: manuscript and cursive. These styles, which constitute the two "levels" of the treatment variable, are to be compared for a population of fourth grade children. The measure of legibility to be employed is based on the number of regressions in the eye movements of adult readers as they read a standard passage written in one or the other of these styles. Previous research with this measure has given rise to an error variance of 10.00, an estimate which may be taken as a population value for this purpose. The completely randomized design is to be used. For a difference of 3.0 between the population means, the experimenter wishes the power of the F-test to equal .90. The .05 level of significance has been adopted

For this situation

$$\lambda = \sqrt{\frac{\Sigma (\mu_i - \mu)^2}{k\sigma^2}} = \sqrt{\frac{(1.5)^2 + (1.5)^2}{2(10)}} = .47.$$

Entering Figure 1 with this value, and using the curve for $P = .90$, $\alpha = .05$, we read the required number of observations per treatment to equal 24+ or 25.

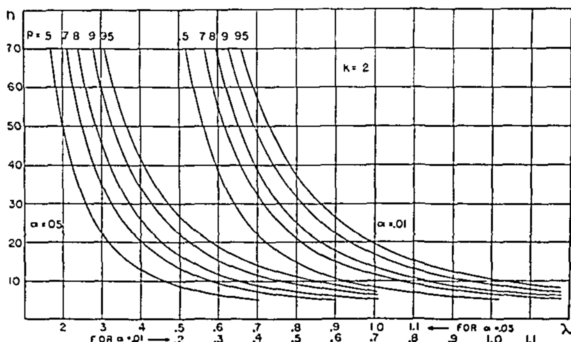


FIG 1 Curves of constant power (P) for the test of main effects with k=2.

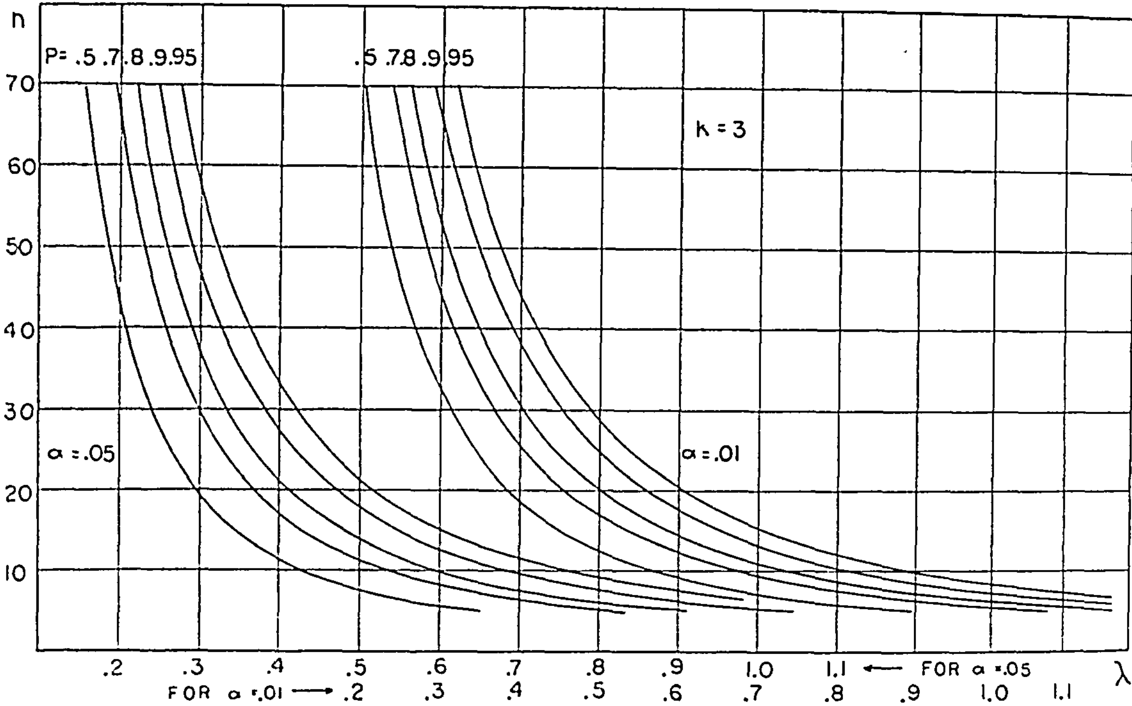


Fig. 2. Curves of constant power (P) for the test of main effects with k=3.

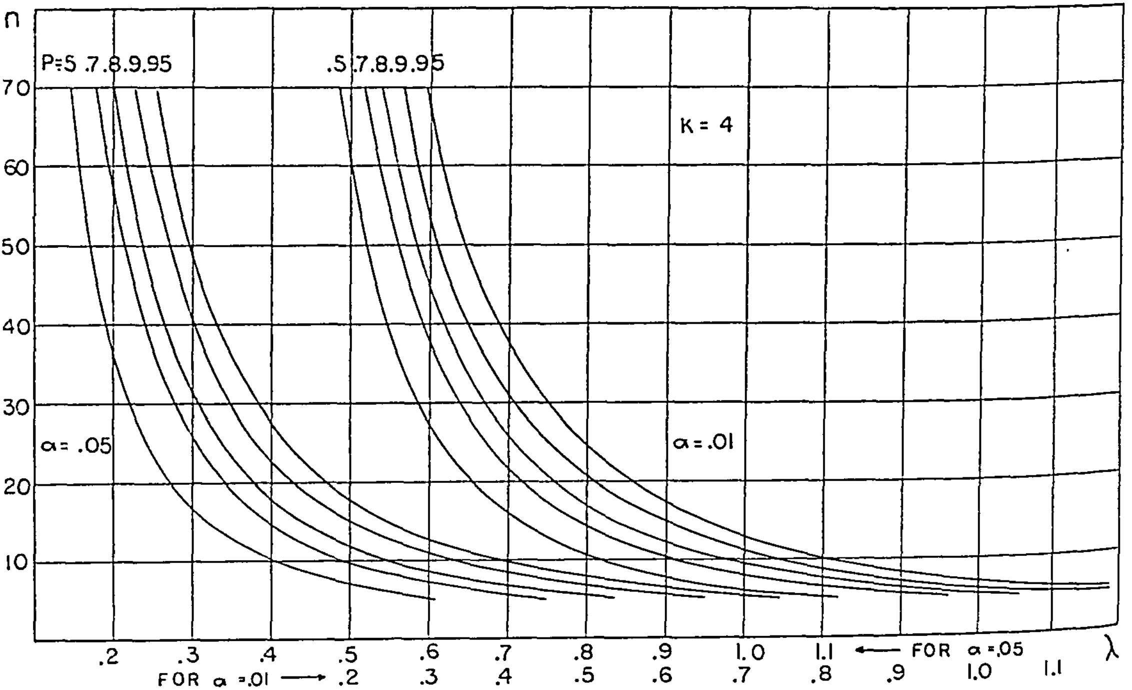


Fig. 3. Curves of constant power (P) for the test of main effects with k=4.

In this example the difference between the population means and the error variance were separately specified. It is often the case, however, that the alternative hypothesis can be defined as a proportion of the error variance. For example, the experimenter might desire a certain power against the alternative

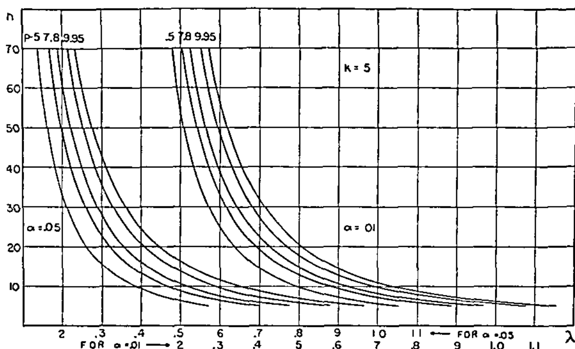


FIG 4. Curves of constant power (P) for the test of main effects with k=5.

$$\frac{\sum (\mu_j - \mu)^2}{k} = .10\sigma^2.$$

In this case

$$\lambda = \sqrt{\frac{\Sigma (\mu_j - \mu)^2}{k\sigma^2}} = \sqrt{.10} = .32.$$

The value of λ is thus specified without an explicit statement of the absolute differences between treatment population means.

6. Note. Steps 2 and 3 in the derivation of these charts are based on the relationship which holds between f_2 and n for the completely randomized design. Since this relationship varies from one experimental design to another, these charts are strictly valid only for the completely randomized setup. For precisely accurate determination of the value of n in any other design, a unique set of charts for that design would be required. Charts for the randomized block design, for example, would be based on the relationship

$$f_2 = (k - 1)(n - 1)$$

or

$$n = 1 + \frac{f_2}{(k - 1)}$$

Charts for the test of the factor with k levels in the $k \times h$ factorial design would be based on the relationship

$$f_2 = k(n - h)$$

or

$$n = h + \frac{f_2}{k}.$$

However, from charts specifically constructed for each of these designs it was found that when $k(n - 1) \geqq 20$ the relationship between λ and n is almost identical for all three designs. Little inaccuracy results from the application of charts based upon the relationship which holds for the completely randomized design.

The relatively small error involved in using the present charts for planning randomized block and factorial experiments is demonstrated by the values in Table 1. In this table the appropriate numbers of observations are indicated for selected experimental conditions involving the three types of designs. The values of n for the randomized block and factorial designs were derived from the charts specially constructed for these designs. It may be seen that in every instance the value of n read from charts constructed for the completely randomized design is only slightly smaller than that read from charts specific to the other designs. The underestimate is less than one observation in almost

TABLE 1

Comparative Values of n for Completely Randomized, Randomized Block, and Factorial Experiments ($\alpha = .05$)

k	P	λ	n		
			Completely Randomized	Randomized Block	$k \times k$ Factorial
2	.5	.525	8.0	8.9	8.2
		.358	16.0	16.9	16.1
		.257	30.0	30.9	30.1
		.181	60.0	60.7	60.0+
	.95	.967	8.0	9.1	8.2
		.657	16.0	17.1	16.1
		.473	30.0	30.9	30.0+
		.333	60.0	60.5	60.0+
4	.5	.450	8.0	8.6	8.8
		.308	16.0	16.3	16.2
		.220	30.0	30.1	30.0+
		.156	60.0	60.0	60.0+
	.95	.770	8.0	8.6	8.8
		.532	16.0	16.5	16.3
		.382	30.0	30.3	30.0+
		.270	60.0	60.0+	60.0

all cases. Therefore, for the practical purpose of approximating the necessary number of observations per treatment in randomized block and factorial experiments, it would seem sufficiently precise to use values read from the present charts.

REFERENCES

[1] A. J. DUNCAN, "Charts of the 10% and 50% points of the operating characteristic curves for fixed effects analysis of variance F tests, $\alpha = .01$ and .05," *J. Amer. Stat. Assoc.*, Vol. 52 (1957), pp. 345–349.
[2] M. FOX, "Charts for the power of the F-test," *Ann. Math. Stat.*, Vol 27 (1956), pp. 484–497.
[3] E. LEHMER, "Inverse tables of probabilities of errors of the second kind," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 388–398.
[4] E. S. PEARSON AND H. O. HARTLEY, "Charts for the power function for analysis of variance tests, derived from the noncentral F-distribution," *Biometrika*, Vol. 38 (1951), pp. 112–130.
[5] P. C TANG, "The power function of the analysis of variance test with tables and illustrations of their use," *Stat. Res. Mem.*, Vol 2 (1938), pp 126–149 + tables.

LIMITING DISTRIBUTIONS IN SOME OCCUPANCY PROBLEMS[1, 2]

By Irving Weiss

Bell Telephone Laboratories, North Andover, Mass.

Section I

Introduction. The classical occupancy problem is concerned with the randc distribution of a specified number of objects (r) in a given number of cells (N No restriction is placed on the number of objects in any cell other than th the total number of objects equals r. The problem of finding exactly m ce empty for the case with r and N finite, and with all arrangements of r objec having equal probability can be expressed in closed form [1]. However, f large N, use of this formula for computation becomes exceedingly tediou Several authors, [2] and [3] have stated without proof that under suitable r strictions on N, r the limiting distribution of the number of unoccupied cel as N, r approach infinity is normal.

By imposing the restriction $\alpha = r/N$, $\alpha > 0$, it will be shown that in tl above occupancy problem the asymptotic distribution of the number of unoc cupied cells is normal.

A modification of the above occupancy problem is the following: q object are randomly distributed among N cells such that no more than one object i in any cell. The procedure is repeated w times. For example, with $w = k$, th maximum number of objects in any cell is k, one for each of k trials. It can b shown that by restricting $qw = \alpha N$, $\alpha > 0$, the normal asymptotic result give above holds. Also, by imposing the restriction $qw = N \log N/\lambda$ the number o unoccupied cells has asymptotically a Poisson distribution. This is an exten sion of the same results listed by Feller [1] for the classical occupancy problem Proofs for the modified occupancy problem have been given by the author [7] and will not be given in this paper.

2. Outline of proof. In showing asymptotic normality our method will em ploy moments. We show that the moments converge to the moments of the normal distribution. From this it follows (by a theorem in Uspensky [4]) that the distribution of our random variable converges uniformly to the normal distribution.

3. Main results. With $\alpha = r/N$, $\alpha > 0$, we define a random variable X_j as follows:

$$X_j = 1 \quad \text{if cell } j \text{ is unoccupied after } r \text{ tosses.}$$
$$= 0 \quad \text{otherwise.}$$

Received January 9, 1956; revised December 26, 1957.

[1] This paper is part of the author's Ph.D. dissertation at Stanford University.

[2] Sponsored by the Office of Naval Research.

Assuming all N events are equally likely and that the r trials are independent of each other:

$$E(X_{i_1} \cdot X_{i_2}, \cdots, X_{i_s}) = \left(1 - \frac{s}{N}\right)^r.$$

Let X equal the number of unoccupied cells

$$X = \sum_{i=1}^{N} X_i$$

$$E(X) = N\left(1 - \frac{1}{N}\right)^r = N\left(1 - \frac{1}{N}\right)^{aN}$$

$$\lim_{N \to \infty} \frac{E(X)}{N} = e^{-a}$$

As N becomes infinite, $E(X)$ becomes infinite but $E(X)/N$ approaches a finite limit.

We will prove that the random variable X has an asymptotically normal distribution by showing that

$$\lim_{N \to \infty} \frac{\mu_k}{(\sigma^2)^{k/2}} = 1 \cdot 3 \cdots (k-1) \quad \text{for } k \text{ even}$$

$$= 0 \quad \text{for } k \text{ odd}$$

The general kth moment, μ_k, is

$$\mu_k = E(X - E(X))^k = \sum_{r=0}^{k} (-1)^r \binom{k}{r} E(X^{k-r})(E(X))^r.$$

As shown in Theorem 1 of Section II, by using Stirling numbers of the and second kind, μ_k can be expressed as follows:

$$\mu_k = \sum_{r=0}^{k} \sum_{p=1}^{k-r} \sum_{j=1}^{p} (-1)^r \binom{k}{r} N^{p+r} S_p^j \mathfrak{S}_{k-r}^{p} \left(1 - \frac{p}{N}\right)^{aN} \left(1 - \frac{1}{N}\right)^{aNr}.$$

It can be shown (see [5]) that

$$\left(1 - \frac{p}{N}\right)^{aN} \left(1 - \frac{1}{N}\right)^{aNr} = exp\left[-a(p + r)\right] exp\left[-\sum_{t=1}^{\infty} \frac{a(r + p^{t+1})}{(t + 1)N^t}\right].$$

Now

$$exp\left[-\sum_{t=1}^{\infty} \frac{a(r + p^{t+1})}{(t + 1)N^t}\right] = \sum_{v=0}^{\infty} \sum_{\{m_j \mid \sum_{j=1}^{v} j m_j = v\}} \frac{q_1^{m_1}}{m_1!} \cdots \frac{q_v^{m_v}}{m_v!} \cdot \frac{1}{N^v}$$

$$= \sum_{v=0}^{\infty} K_v(r, s) \frac{1}{N^v}$$

where

$$a_i = \frac{-(r + p^{i+1})}{i + 1}$$

Substituting above and noting that $S_p^0 = S_{k-r}^0 = 0$ we have

$$\mu_k = \sum_{r=0}^{k} \sum_{s=r}^{k} \sum_{v=r}^{s} (-1)^r \binom{k}{r} S_{s-r}^{v-r} S_{k-r}^{s-r} N^v e^{-\alpha s} \sum_{n=0}^{\infty} K_n(r, s) \frac{1}{N^n}$$

where

$$p + r = s$$
$$j + r = v$$

Collecting like powers of N

$$\mu_k = \sum_{v=k}^{-\infty} N^v \left[\sum_{s=v}^{k} e^{-\alpha s} b_{s,v,0} + \sum_{s=v+1}^{k} e^{-\alpha s} b_{s,v+1,1} + \cdots + e^{-\alpha k} b_{k,k,k-v} \right]$$

(1) $$= \sum_{v=k}^{-\infty} N^v \left[\sum_{s=v}^{k} e^{-\alpha s} \left(\sum_{n=0}^{s-v} b_{s,v+n,n} \right) \right]$$

$$= \sum_{v=k}^{[k/2]} N^v \left[\sum_{s=v}^{k} e^{-\alpha s} a_s(v) \right] + R(r, N)$$

where

$$b_{s,v+n,n} = \sum_{r=0}^{k} (-1)^r \binom{k}{r} S_{s-r}^{v+n-r} S_{k-r}^{s-r} K_n(r, s)$$
$$= \sum_{r=0}^{k} (-1)^r \binom{k}{r} f(k - r)$$
$$= \Delta^k f(0)$$

and

$$[k/2] = k/2 \qquad \text{for } k \text{ even}$$
$$= \frac{k-1}{2} \qquad \text{for } k \text{ odd}$$

As shown in [6] $b_{s,v+n,n}$ is the kth difference of $f(0)$. By Lemma 1,

$$b_{s,v+n,n} \equiv 0 \qquad \text{for} \qquad v > k/2$$
$$= ck! \qquad \text{for} \qquad v = k/2$$

where c is the product of the coefficients of the highest degree terms in r o
S_{s-r}^{v+n-r}, S_{k-r}^{s-r} and $K_n(r, s)$.

For a given k, $R(r, N)$ is a bounded function of r and N. This is an immediat consequence of the analyticity of μ_k. From (1), the highest power of N i

$R(r, N)$ is $N^{[k/2]-1}$. Therefore

$$R(r, N) = O(N^{[k/2]-1}).$$

Incorporating these results in (1) for k even

$$\mu_k = N^{k/2} \sum_{s=k/2}^{k} e^{-\alpha s} a_s(k) + O(N^{k/2-1})$$

where

$$a_s(k) = \sum_{n=0}^{s-k/2} b_{s,k/2+n,n} = \sum_{n=0}^{s-k/2} ck!$$

Using Lemma 1, it follows that

$$a_s(k) = D_{k/2,0}(-1)^h(\alpha + 1)^h \binom{k/2}{h}$$

where

$$D_{k/2,0} = 1 \cdot 3 \cdots (k - 1)$$

$$h = s - k/2$$

Substituting above

$$\mu_k = N^{k/2}(e^{-\alpha} - (\alpha + 1)e^{-2\alpha})^{k/2} D_{k/2,0} + O(N^{k/2-1})$$

Noting that $\sigma^2 = \mu_2$, forming the ratio

$$\frac{\mu_k}{(\sigma^2)^{k/2}} = \frac{D_{k/2,0} N^{k/2}(e^{-\alpha} - (\alpha + 1)e^{-2\alpha})^{k/2} + O(N^{k/2-1})}{N^{k/2}(e^{-\alpha} - (\alpha + 1)e^{-2\alpha})^{k/2} + O(N^{k/2-1})}$$

dividing numerator and denominator by $N^{k/2}$ and then letting $N \to \infty$,

$$\lim_{N\to\infty} \frac{\mu_k}{(\sigma^2)^{k/2}} = D_{k/2,0} \quad \text{for} \quad k = 2, 4, \cdots.$$

For k an odd positive integer,

$$\lim_{N\to\infty} \frac{\mu_k}{(\sigma^2)^{k/2}} = 0.$$

This follows from the fact that $b_{s,v+n,n} = 0$ for $v \geqq k/2$ as v being a positive integer cannot equal $k/2$. Therefore,

$$\mu_k = O(N^{(k-1)/2})$$

while

$$(\sigma^2)^{k/2} = O(N^{k/2}).$$

SECTION II

THEOREM 1.

$$\mu_k = \sum_{r=0}^{k} \sum_{p=1}^{k-r} \sum_{j=1}^{p} (-1)^r \binom{k}{r} N^{j+r} S_p^j \mathfrak{S}_{k-r}^p \left(1 - \frac{p}{N}\right)^{\alpha N} \left(1 - \frac{1}{N}\right)^{\alpha N r}$$

where S_p^j *and* $\mathfrak{S}_{k-r}^p$ *are Stirling numbers of the first and second kind respectively.*

PROOF.

$$(2) \qquad \mu_k = E(X - E(X))^k = \sum_{r=0}^{k} (-1)^r \binom{k}{r} E(X^{k-r})[E(X)]^r.$$

By the multinomial expansion with

$$\lambda(s_i) = 1 \quad \text{for} \quad s_i > 0$$
$$= 0 \quad \text{for} \quad s_i = 0$$

and

$$\sum_{i=1}^{N} \lambda(s_i) = p \qquad 1 \leqq p \leqq s$$
$$X_i^{s_i} = X_i^{\lambda(s_i)} \qquad \text{as } X_i = 0 \text{ or } 1.$$

we have

$$X^s = \sum_{p=1}^{s} \sum_{\left\{\substack{s_i : \sum_{i=1}^N \lambda(s_i) = p \\ \sum_{i=1}^N s_i = s}\right\}} \frac{s!}{s_1! \cdots s_N!} \binom{N}{p} X_1^{\lambda(s_1)} \cdots X_N^{\lambda(s_N)}$$

From Jordan [5]

$$\mathfrak{S}_s^p = \frac{s!}{p!} \sum_{\left\{\substack{s_i : s_i > 0 \\ \sum_{i=1}^p s_i = s}\right\}} \frac{1}{s_1! \cdots s_p!}$$

Substituting above to eliminate the second summation and taking expectations,

$$E(X^s) = \sum_{p=1}^{s} p! \binom{N}{p} \mathfrak{S}_s^p E(X_1^{\lambda(s_1)} \cdots X_N^{\lambda(s_N)})$$
$$= \sum_{p=1}^{s} (N)_p \left(1 - \frac{p}{N}\right)^{\alpha N} \mathfrak{S}_s^p$$

From Jordan [5]

$$(N)_p = \sum_{j=1}^{p} S_p^j N^j$$

Substituting above in (2) with

$$[E(X)]^r = N^r\left(1 - \frac{1}{N}\right)^{\alpha Nr}$$

yields the desired result.

THEOREM 2. *The degree of $K_n(r, s)$ defined in equation (1), considered as a polynomial in r, is obtained from the term of the summation in which $m_1 = n$ and $m_2 = m_3 = \cdots = m_n = 0$.*

PROOF. The highest power of r in a_i is $i + 1$. For a given n we have to determine $m_1, \cdots, m_n$ which will maximize the highest power of r subject to the restriction that

$$\sum_{i=1}^{n} im_i = n.$$

Maximizing the power of r is equivalent to maximizing

$$2m_1 + 3m_2 + \cdots + (n + 1)m_n = n + \sum_{i=1}^{n} m_i$$

Maximizing $\sum_{i=1}^{n} m_i$ subject to the above restraint yields

$$\sum_{i=1}^{n} m_i = n - \sum_{i=2}^{n} (i - 1)m_i$$

The maximum is attained when $m_i = 0; i = 2, \cdots, n$. Therefore, the power of r is maximized when $m_1 = n$. From the definition of a_1, it is readily seen that the degree of r in $K_n(r, s)$ is $2n$ and that the coefficient of this highest degree term is $(-\alpha)^n/2^n n!$.

LEMMA 1. *Let $b_{s,v+n,n}$ be defined as above. Then*

$$\begin{aligned} b_{s,v+n,n} &\equiv 0 \quad \text{for} \quad v > k/2 \\ &= ck! \quad \text{for} \quad v = k/2 \end{aligned}$$

where

$$c = \frac{C_{(s-k/2-n),0}}{[2(s - k/2 - n)]!} \cdot \frac{D_{(k-s),0}}{[2(k - s)]!} \cdot \frac{(-\alpha)^n}{n!2^n}.$$

PROOF. From Jordan [5], S_n^{n-m} and $\mathfrak{S}_n^{n-m}$ are polynomials in n of degree $2m$, i e.:

$$S_n^{n-m} = C_{m,0}\frac{(n)_{2m}}{(2m)!} + \text{terms in } n \text{ of degrees less than } 2m$$

$$\mathfrak{S}_n^{n-m} = D_{m,0}\frac{(n)_{2m}}{(2m)!} + \text{terms in } n \text{ of degrees less than } 2m$$

where

$$C_{m,0} = (-1)^m D_{m,0}$$

As the product of a finite number of polynomials is also a polynomial,

$$S_{s-r}^{v+n-r} S_{k-r}^{s-r} K_n(r, s)$$

is also a polynomial. Its degree in r for fixed v, s, n is

$$2(s - v - n) + 2(k - s) + 2n = 2(k - v) \cdot$$

It follows from elementary properties from the calculus of finite differences that

$$\begin{aligned} b_{s,v+n,n} &\equiv 0 \qquad \text{for} \qquad v > k/2 \\ &= ck! \qquad \text{for} \qquad v = k/2 \end{aligned}$$

where c is the coefficient of r^k in the product polynomial. That c is the product of the above three factors is apparent from the polynomial expansion of Stirling numbers and from Theorem 2.

Acknowledgment. I am deeply indebted to Dr. Joseph F. Daly whose helpful guidance made this work possible.

REFERENCES

[1] WILLIAM, FELLER, "An introduction to probability theory and its applications," John Wiley & Sons, New York, 1950.
[2] F. N. DAVID, "Two combinatorial tests of whether a sample has come from a given population," *Biometrika*, 37, 1950, pp. 97–110.
[3] V. W. SCHAFER, "Das Mutungsproblem der Besetzungs-Verteilung," *Mitteilungsblatt Math. Statist.* 5, (1954), 1–38.
[4] J. V. USPENSKY, "Introduction to mathematical probability," McGraw-Hill Book Co., New York, 1937.
[5] CHARLES JORDAN, "Calculus of Finite Differences," Chelsea Publishing Co., New York, 1950.
[6] TOMLINSON FORT, "Finite differences and difference equations in the real domain," Oxford, 1948.
[7] IRVING WEISS, "Limiting distributions in some occupancy problems," Ph.D. dissertation, Stanford University, Stanford, California.

AN EXTENSION OF BOX'S RESULTS ON THE USE OF THE F DISTRIBUTION IN MULTIVARIATE ANALYSIS

By Seymour Geisser and Samuel W. Greenhouse

National Institute of Mental Health

1. Introduction and summary. The mixed model in a 2-way analysis of variance is characterized by a fixed classification, e.g., treatments, and a random classification, e.g., plots or individuals. If we consider k different treatments each applied to everyone of n individuals, and assume the usual analysis of variance assumptions of uncorrelated errors, equal variances and normality, an appropriate analysis for the set of nk observations x_{ij}, $i = 1, 2, \cdots n$, $j = 1, 2 \cdots k$, is

Source	*D.F.*	*F*
Treatments	$k - 1$	$\dfrac{\text{mean square for treatments}}{\text{mean square for } T \times I}$
Individuals	$n - 1$	
Treat. $\times$ Ind.	$(k - 1)(n - 1)$	

where the F ratio under the null hypothesis has the F distribution with $(k - 1)$ and $(k - 1)(n - 1)$ degrees of freedom. As is well known, if we extend the situation so that the errors have equal correlations instead of being uncorrelated, the F ratio has the same distribution. Under the null hypothesis, the numerator estimates the same quantity as the denominator, namely, $(1 - \rho)\sigma^2$, where ρ is the constant correlation coefficient among the treatments. This case can also be considered as a sampling of n vectors (individuals) from a k-variate normal population with variance-covariance matrix

$$V = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & & & \vdots \\ \vdots & & & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}.$$

If we consider this type of formulation and suppose the k treatment errors to have a multivariate normal distribution with unknown variance-covariance matrix (the same for each individual), then the usual test described above is valid for $k = 2$. For $k > 2$, and $n \geqq k$, Hotelling's T^2 is the appropriate test for the homogeneity of the treatment means. However, the working statistician is sometimes confronted with the case where $k > n$, or he does not have the adequate means for computing large order inverse matrices and would therefore like to use the original test ratio which in general does not have the requisite F distribution. Box [1] and [2] has given an approximate distribution of the test ratio to be $F[(k - 1)\epsilon, (k - 1)(n - 1)\epsilon]$ where ϵ is a function of the popula-

Received May 27, 1957; revised January 29, 1958.

tion variances and covariances and may further be approximated by the sample variances and covariances. We show in Section 3 that $\epsilon \geqq (k-1)^{-1}$, and therefore a conservative test would be $F(1, n-1)$.

Box referred only to one group of n individuals. We shall extend his results to a frequently occurring case, namely, the analysis of g groups where the αth group has n_α individuals, $\alpha = 1, 2, \cdots g$, and $\sum_{\alpha=1}^{g} n_\alpha = N$. We will show that the treatment mean square and the treatment $\times$ group interaction can be tested in the same approximate fashion by using the Box procedure.

2. Extension to g groups. Consider a mixed model, k treatments, each applied to N individuals where the N individuals are subdivided into g groups so that we have chosen a random sample of n_α individuals from the αth group. The observations are $x_{ij\alpha}$, $i = 1, \cdots n_\alpha$, $j = 1, \cdots, k$, $\alpha = 1, \cdots, g$ and

$$\sum_{\alpha=1}^{g} n_\alpha = N.$$

Therefore we get the following array for the αth group

$$\begin{matrix} x_{11\alpha} & \cdots & x_{1k\alpha} \\ \vdots & & \vdots \\ x_{n_\alpha 1\alpha} & \cdots & x_{n_\alpha k\alpha} \end{matrix}$$

We may consider the joint distribution of the $x_{ij\alpha}$ to be represented by the vector variable

$$x' = (x_{111} \cdots x_{1k1} \cdots x_{n_1 11} \cdots x_{n_1 k1} \cdots x_{11g} \cdots x_{1kg} \cdots x_{n_g 1g} \cdots x_{n_g kg})$$

where $Ex' = \mu'$ and x' has a kN multivariate normal distribution with variance-covariance matrix

$$\Lambda = \begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & V_g \end{pmatrix}$$

and

$$V_\alpha = \begin{pmatrix} V & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & V \end{pmatrix},$$

where V is a matrix of order k, V_α is of order kn_α and Λ is of order kN.

Let $Ex_{ij\alpha} = \mu_{j\alpha}$ and

$$N^{-1} \sum_{\alpha=1}^{g} n_\alpha \mu_{j\alpha} = \mu_{j.} \text{ is the mean of the } j\text{th treatment,}$$

$$k^{-1} \sum_{j=1}^{k} \mu_{j\alpha} = \mu_{.\alpha} \text{ is the mean of the } \alpha\text{th group, and}$$

$$k^{-1} \sum_{j=1}^{k} \mu_{j.} = N^{-1} \sum_{\alpha=1}^{g} n_\alpha \mu_{.\alpha} = \mu_{..} \text{ the grand mean.}$$

TABLE 1

Source	D.F.	S.S.	F
Treatments	$k-1$	Q_1	$F_1 = (N-g)Q_1/Q_5$
Groups	$g-1$	Q_2	$F_2 = (N-g)Q_2/(g-1)Q_3$
Ind. Within Groups	$N-g$	Q_3	
Treat. × Groups	$(k-1)(g-1)$	Q_4	$F_3 = (N-g)Q_4/(g-1)Q_5$
Treat. × Ind. Within Groups.	$(k-1)(N-g)$	Q_5	
Total	$Nk-1$		

We will now partition the total sum of squares into 5 constituent sums of squares, as one would usually do with a mixed model that satisfied all the usual analysis of variance assumptions.

Let S be defined as the matrix of the quadratic form representing the correction factor which is the square of the grand total of all the observations divided by kN. S is a $kN \times kN$ matrix whose elements are all $(kN)^{-1}$. Further let a matrix M of sub-matrices $M_{\alpha\beta}$ be denoted as

$$\{M_{\alpha\beta}\} = \begin{pmatrix} M_{11} & \cdots & M_{1g} \\ \vdots & & \vdots \\ M_{g1} & \cdots & M_{gg} \end{pmatrix}.$$

If $M_{\alpha\beta} = 0$, for $\alpha \neq \beta$, let the resulting matrix be denoted by

$$\{M_\alpha\} = \begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & M_g \end{pmatrix}.$$

Now let

$$Q_1 = x'Ax = N \sum_{j=1}^{k} (\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot})^2$$

and

$$A = \{N^{-1}A_{\alpha\beta}\} - S,$$

where $A_{\alpha\beta}$ is the matrix of $n_\alpha \times n_\beta$ matrices each of which is the $k \times k$ identity matrix.

Let

$$Q_2 = x'Bx = k \sum_{\alpha=1}^{g} n_\alpha (\bar{x}_{\cdot\cdot\alpha} - \bar{x}_{\cdot\cdot\cdot})^2,$$

where $B = \{n_\alpha^{-1}B_\alpha\} - S$ and B_α is the matrix of $n_\alpha \times n_\alpha$ matrices each of which is of order $k \times k$, and is of the form

$$E = k^{-1}1_k 1_k',$$

where $1_k' = (1, \cdots, 1)$.

tion variances and covariances and may further be approximated by the sample variances and covariances. We show in Section 3 that $\epsilon \geqq (k - 1)^{-1}$, and therefore a conservative test would be $F(1, n - 1)$.

Box referred only to one group of n individuals. We shall extend his results to a frequently occurring case, namely, the analysis of g groups where the αth group has n_α individuals, $\alpha = 1, 2, \cdots g$, and $\sum_{\alpha=1}^{g} n_\alpha = N$. We will show that the treatment mean square and the treatment $\times$ group interaction can be tested in the same approximate fashion by using the Box procedure.

2. Extension to g groups. Consider a mixed model, k treatments, each applied to N individuals where the N individuals are subdivided into g groups so that we have chosen a random sample of n_α individuals from the αth group. The observations are $x_{ij\alpha}$, $i = 1, \cdots n_\alpha$, $j = 1, \cdots, k$, $\alpha = 1, \cdots, g$ and

$$\sum_{\alpha=1}^{g} n_\alpha = N.$$

Therefore we get the following array for the αth group

$$\begin{matrix} x_{11\alpha} & \cdots & x_{1k\alpha} \\ \vdots & & \vdots \\ x_{n_\alpha 1\alpha} & \cdots & x_{n_\alpha k\alpha} \end{matrix}$$

We may consider the joint distribution of the $x_{ij\alpha}$ to be represented by the vector variable

$$x' = (x_{111} \cdots x_{1k1} \cdots x_{n_1 11} \cdots x_{n_1 k1} \cdots x_{11g} \cdots x_{1kg} \cdots x_{n_g 1g} \cdots x_{n_g kg})$$

where $Ex' = \mu'$ and x' has a kN multivariate normal distribution with variance-covariance matrix

$$\Lambda = \begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & V_g \end{pmatrix}$$

and

$$V_\alpha = \begin{pmatrix} V & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & V \end{pmatrix},$$

where V is a matrix of order k, V_α is of order kn_α and Λ is of order kN.

Let $Ex_{ij\alpha} = \mu_{j\alpha}$ and

$$N^{-1} \sum_{\alpha=1}^{g} n_\alpha \mu_{j\alpha} = \mu_{j.} \text{ is the mean of the } j\text{th treatment,}$$

$$k^{-1} \sum_{j=1}^{k} \mu_{j\alpha} = \mu_{.\alpha} \text{ is the mean of the } \alpha\text{th group, and}$$

$$k^{-1} \sum_{j=1}^{k} \mu_{j\cdot} = N^{-1} \sum_{\alpha=1}^{g} n_\alpha \mu_{\cdot\alpha} = \mu_{\cdot\cdot} \text{ the grand mean.}$$

TABLE 1

Source	D.F.	S.S	F
Treatments	$k - 1$	Q_1	$F_1 = (N - g)Q_1/Q_5$
Groups	$g - 1$	Q_2	$F_2 = (N - g)Q_2/(g - 1)Q_3$
Ind. Within Groups	$N - g$	Q_3	
Treat. × Groups	$(k - 1)(g - 1)$	Q_4	$F_3 = (N - g)Q_4/(g - 1)Q_5$
Treat. × Ind. Within Groups.	$(k - 1)(N - g)$	Q_5	
Total	$Nk - 1$		

We will now partition the total sum of squares into 5 constituent sums of squares, as one would usually do with a mixed model that satisfied all the usual analysis of variance assumptions.

Let S be defined as the matrix of the quadratic form representing the correction factor which is the square of the grand total of all the observations divided by kN. S is a $kN \times kN$ matrix whose elements are all $(kN)^{-1}$. Further let a matrix M of sub-matrices $M_{\alpha\beta}$ be denoted as

$$\{M_{\alpha\beta}\} = \begin{pmatrix} M_{11} & \cdots & M_{1g} \\ \vdots & & \vdots \\ M_{g1} & \cdots & M_{gg} \end{pmatrix}.$$

If $M_{\alpha\beta} = 0$, for $\alpha \neq \beta$, let the resulting matrix be denoted by

$$\{M_\alpha\} = \begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & M_g \end{pmatrix}.$$

Now let

$$Q_1 = x'Ax = N \sum_{j=1}^{k} (\bar{x}_{\cdot j \cdot} - \bar{x}_{\cdot\cdot\cdot})^2$$

and

$$A = \{N^{-1}A_{\alpha\beta}\} - S,$$

where $A_{\alpha\beta}$ is the matrix of $n_\alpha \times n_\beta$ matrices each of which is the $k \times k$ identity matrix.

Let

$$Q_2 = x'Bx = k \sum_{\alpha=1}^{g} n_\alpha (\bar{x}_{\cdot\cdot\alpha} - \bar{x}_{\cdot\cdot\cdot})^2,$$

where $B = \{n_\alpha^{-1}B_\alpha\} - S$ and B_α is the matrix of $n_\alpha \times n_\alpha$ matrices each which is of order $k \times k$, and is of the form

$$E = k^{-1}\mathbf{1}_k\mathbf{1}_k',$$

where $\mathbf{1}_k' = (1, \cdots, 1)$.

Let

$$Q_3 = x'Cx = k \sum_{\alpha=1}^{g} \sum_{i=1}^{n_\alpha} (\bar{x}_{i\cdot\alpha} - \bar{x}_{\cdot\cdot\alpha})^2,$$

where $C = \{n_\alpha^{-1} C_\alpha\}$ and $C_\alpha = n_\alpha E_\alpha - B_\alpha$, where

$$E_\alpha = \begin{pmatrix} E & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & E \end{pmatrix}.$$

Let

$$Q_4 = x'Dx = \sum_{\alpha=1}^{g} n_\alpha \sum_{j=1}^{k} (\bar{x}_{\cdot j\alpha} - \bar{x}_{\cdot\cdot\alpha} - \bar{x}_{\cdot j\cdot} + \bar{x}_{\cdot\cdot\cdot})^2,$$

where

$$D = \{n_\alpha^{-1}(A_\alpha - B_\alpha)\} + \{N^{-1}(B_{\alpha\beta} - A_{\alpha\beta})\},$$

and here $M_{\alpha\beta}$ is a matrix of $n_\alpha \times n_\beta$ matrices each of order k where here refers to the matrix of identity matrices and $B_{\alpha\beta}$ refers to matrices of E's.

Let

$$Q_5 = x'Fx = \sum_{\alpha=1}^{g} \sum_{j=1}^{k} \sum_{i=1}^{n_\alpha} (x_{ij\alpha} - \bar{x}_{\cdot j\alpha} - \bar{x}_{i\cdot\alpha} + \bar{x}_{\cdot\cdot\alpha})^2,$$

where $F = \{n_\alpha^{-1} F_\alpha\}$ and $F_\alpha = n_\alpha(I_\alpha - E_\alpha) + B_\alpha - A_\alpha$, where

$$I_\alpha - E_\alpha = \begin{pmatrix} I - E & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & I - E \end{pmatrix}.$$

Now it is easy to show even for arbitrary V, the basic matrix of Λ, that

$$A\Lambda F = D\Lambda F = B\Lambda C = B\Lambda F = C\Lambda D = 0.$$

Hence by a result due to Carpenter [3], Q_1 and Q_4 are independent of Q_5 is independent of Q_3 and Q_5, and Q_3 is independent of Q_4. Further as Box shown if $Q = (x - \mu)'M(x - \mu)$ where x' has variance-covariance matrix the sth cumulant of Q, $K_s(Q) = 2^{s-1}(s-1)! \operatorname{Tr} (\Lambda M)^s$ where Tr stands for trace of a matrix. Hence by straightforward algebra we get

$$K_1(Q_1) = \operatorname{Tr} V - \operatorname{Tr} EV + N \sum_{j=1}^{k} (\mu_{j.} - \mu_{..})^2,$$

$$K_1(Q_2) = (g - 1) \operatorname{Tr} EV + \sum_{\alpha=1}^{g} n_\alpha (\mu_{\cdot\alpha} - \mu_{\cdot\cdot})^2,$$

$$K_1(Q_3) = (N - g) \operatorname{Tr} EV,$$

$$K_1(Q_4) = (g - 1)(\operatorname{Tr} V - \operatorname{Tr} EV) + \sum_{\alpha=1}^{g} n_\alpha \sum_{j=1}^{k} (\mu_{j\alpha} - \mu_{j\cdot} - \mu_{\cdot\alpha} + \mu_{\cdot\cdot})^2,$$

$K_1(Q_5) = (N - g)(\text{Tr } V - \text{Tr } EV),$

and

$$K_2(Q_1) = 2 \text{ Tr } (\Lambda A)^2 = 2 \text{ Tr } (V - EV)^2 \text{ if } \mu_{j.} = \mu \;,$$
$$K_2(Q_2) = 2 \text{ Tr } (\Lambda B)^2 = 2(g - 1) \text{ Tr } (EV)^2 \text{ if } \mu_{.\alpha} = \mu \;,$$
$$K_2(Q_3) = 2 \text{ Tr } (\Lambda C)^2 = 2(N - g) \text{ Tr } (EV)^2,$$
$$K_2(Q_4) = 2 \text{ Tr } (\Lambda D)^2 = 2(g - 1) \text{ Tr } (V - EV)^2 \text{ if } \mu_{j\alpha} - \mu_{j.} - \mu_{.\alpha} + \mu_{..} = 0,$$
$$K_2(Q_5) = 2 \text{ Tr } (\Lambda F)^2 = 2(N - g) \text{ Tr } (V - EV)^2.$$

From the first cumulants it is clear that under the null hypothesis of no treatment differences, the Expected Mean Square (E.M.S.) for $(k - 1)^{-1}Q_1$ is $(k - 1)^{-1}$ $(\text{Tr } V - \text{Tr } EV)$; under the null hypothesis of no group $\times$ treatment interaction, the E.M.S. of $(g - 1)^{-1}(k - 1)^{-1}Q_4$ is $(k - 1)^{-1}(\text{Tr } V - \text{Tr } EV)$, while the E.M.S. of $(N - g)^{-1}(k - 1)^{-1}Q_5$ is just $(k - 1)^{-1}(\text{Tr } V - \text{Tr } EV)$. Hence under the hypothesis that the treatment means are equal, the numerator and denominator of F_1 estimate the same quantity; and under the hypothesis of no interaction, the numerator and denominator of F_3 estimate the same quantity. Similarly under the hypothesis of no group differences, the numerator and denominator of F_2 estimate the same quantity.

Now using the results of Box ([1], Theorem 6.1) on the approximate distribution of linear sums of chi-square variates, it is clear that F_1 is approximately distributed like $F[(k - 1)\epsilon, (k - 1)(N - g)\epsilon]$ and F_3 is approximately like $F[(g - 1)(k - 1)\epsilon, (k - 1)(N - g)\epsilon]$ while it is obvious that F_2 is exactly distributed like $F(g - 1, N - g)$, where (Box [2])

$$\epsilon = k^2(\bar{v}_{tt} - \bar{v}_{..})^2/(k - 1)\left(\sum_{t=1}^{k}\sum_{s=1}^{k} v_{ts}^2 - 2k\sum_{t=1}^{k}\bar{v}_{t.}^2 + k^2\bar{v}_{..}^2\right)$$

and v_{ts} are the elements of V, $\bar{v}_{tt}$ is the mean of the diagonal terms, $\bar{v}_{t.}$ is the mean of the tth row (or tth column) and $\bar{v}_{..}$ is the grand mean. This result is easily extended to the fixed interactions in an r-way classification where one of the ways is individuals divided into g-groups and the other $r - 1$ classifications are fixed.

3. A lower bound on ϵ. Clearly, the formulation of the degrees of freedom with which we enter the F-table requires the computation of the elements of the variance-covariance matrix. We now present a lower limit on ϵ independent of these elements. This limit, although obvious and simple, may be too conservative.

From Theorem 6.1 Box [1], it is easy to show that

$$\epsilon = (k - 1)^{-1}[\text{Tr } (V - EV)]^2 / \text{Tr } (V - EV)^2,$$
$$\epsilon = (k - 1)^{-1}\left(\sum_{j=1}^{k}\lambda_j\right)^2 \Big/ \sum_{j=1}^{k}\lambda_j^2,$$

where $\lambda_j (j = 1 \cdots k)$ are the latent roots of $(V - EV)$ and are non-negative. But $(\sum \lambda_j)^2 \geqq \sum \lambda_j^2$. Therefore $\epsilon \geqq (k - 1)^{-1}$. Hence, F_1 is conservatively

distributed like $F(1, N - g)$ and F_3 is conservatively distributed like $F(g - 1, N - g)$. We also note that if $V = \sigma^2 I$ (the usual analysis of variance assumption) all the roots of $V - EV$ are equal except for one which is equal to zero so that $\epsilon = 1$ in this case.

4. A joint test of groups and treatment × group interaction. In psychological problems it is sometimes necessary to test whether several groups form one cluster. This is equivalent to testing jointly groups and group × treatment interaction. The proposed test here is

$$F_0 = (N - g)Q' / (g - 1)Q,$$

where

$$Q' = Q_2 + Q_4 \qquad \text{and} \qquad Q = Q_3 + Q_5 .$$

It is clear from Section 2 that the numerator and denominator are independent and

$$K_1(Q') = (g - 1) \operatorname{Tr} V + \sum_{\alpha=1}^{g} n_\alpha (\mu_{\cdot\alpha} - \mu_{\cdot\cdot})^2 + \sum_{\alpha=1}^{g} n_\alpha \cdot \sum_{j=1}^{k} (\mu_{j\alpha} - \mu_{j\cdot} - \mu_{\cdot\alpha} + \mu_{\cdot\cdot})^2, \quad K_1(Q) = (N - g) \operatorname{Tr} V;$$

and if $\mu_{\cdot\alpha} = \mu_{\cdot\cdot}$, $\mu_{j\alpha} - \mu_{j\cdot} - \mu_{\cdot\alpha} + \mu_{\cdot\cdot} = 0$, then

$$K_2(Q') = 2(g - 1) \operatorname{Tr} V^2,$$

$$K_2(Q) = 2(N - g) \operatorname{Tr} V^2,$$

and again by using (Theorem 6.1 [1]), F_0 is approximately distributed like $F[(g - 1)k\epsilon', (N - g)k\epsilon']$, where

$$\epsilon' = k\bar{v}_{tt}^2 \Big/ \sum_t^k \sum_s^k v_{ts}^2 .$$

Further it is easy to show that $\epsilon' \geq k^{-1}$ independent of the population variances and covariances and a conservative test would be $F(g - 1, N - g)$. The rationale for this test is that the numerator and denominator of F_0 estimate the same quantity under the null hypothesis of no group effects and no treatment × group effects.

It is of interest to point out and make more explicit the relationship between the foregoing discussion and the general hypothesis in multivariate analysis of the equality of vector means among g populations where all the variables are measured in the same metric. This latter is

$$H_0 \, (\mu_1 = \mu_2 = \cdots = \mu_g),$$

where $\mu_\alpha' = (\mu_{1\alpha}, \mu_{2\alpha}, \cdots, \mu_{k\alpha})$ is the vector mean of the αth group (i.e., multivariate normal population). But the joint test on groups and group × treatment interaction just presented is in effect also a test for the equality of the g

vector means, since the joint null hypothesis of no interaction and equal group means is equivalent to

$$\mu_{j\alpha} - \mu_{j.} - \mu_{.\alpha} + \mu_{.} = 0 \qquad \text{for all } j, \alpha,$$

$$\mu_{.\alpha} = \mu_{.} \qquad \text{for all } \alpha,$$

which is easily seen to imply $\mu_{j\alpha} = \mu_{j.}$ for all α, which is equivalent to $\mu_1 = \mu_2 = \cdots = \mu_g$. Therefore, if the variance-covariance matrices in the g groups can be assumed equal, an approximate test on the hypothesis of equal vector means in multivariate analysis is the F_0 test with ϵ approximated from the sample variances and covariances. It is clear that the conservative F-test which is independent of ϵ can also be used in this case. Furthermore we shall show that if the variance-covariance matrices are not assumed equal, the conservative F-test can be used with the restriction that $n_\alpha = n$.

5. Remarks on unequal variance-covariance matrices. One of the basic assumptions was that each of the N individuals had the same variance-covariance matrix. However if $n_\alpha = n$ for $\alpha = 1, \cdots g$, then we need only assume that individuals in the same group have the same variance-covariance matrix while these variance-covariance matrices may vary from group to group. In this case we get unbiased numerators and denominators of the test ratios as before and the same approximate distributions can be derived, but now the numerator and denominator degrees of freedom have different adjustment factors, each depending upon the different covariance matrices. However it can be easily shown that the lower bounds on these ϵ's are such that F_0, F_1, F_2, and F_3 all have the *same conservative F-test, namely,* $F(1, n - 1)$.

6. Acknowledgment. We are indebted to the referee for detecting an error and for several suggestions which have improved the exposition.

REFERENCES

[1] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification" *Ann. Math. Stat.*, Vol. 25 (1954), pp. 290–302.

[2] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and correlation between errors in the two-way classification," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 484–498.

[3] O. Carpenter, "Note on the extension of Craig's theorem to non-central variates," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 455–457.

A PROPERTY OF ADDITIVELY CLOSED FAMILIES OF DISTRIBUTIONS

By Edwin L. Crow

Boulder Laboratories, National Bureau of Standards

1. Introduction. The property that a linear combination of independent χ^2 variables with coefficients other than unity (or zero) is not distributed as χ^2 has for long been tacitly understood or explicitly stated in studies of the distribution of quadratic forms, the Behrens-Fisher problem, and the precision of estimates of variance components, and in the derivation of tests for the analysis of variance of unbalanced designs. The earliest explicit statement known to the author occurs as a special case of a corollary given without proof by Cochran [2]. A proof depending on the form of the moment-generating function of χ^2 was given by James [6]. The purpose of this note is to state and prove the analogous property for a general class of closed families of distributions, on the basis of work by Teicher [8].

Definition. A one-parameter family of univariate cumulative distribution functions $F(x; \lambda)$ is *additively closed*, if, for any two members $F(x; \lambda_1)$ and $F(x; \lambda_2)$, $F(x; \lambda_1) * F(x; \lambda_2) \stackrel{x}{\equiv} F(x; \lambda_1 + \lambda_2)$.

2. Principal Result.

Theorem. *Consider a one-parameter additively closed family of univariate cumulative distribution functions $F(x; \lambda)$, where λ is* (i) *any positive integer,* (ii) *any positive rational, or* (iii) *any positive real number (except that in case* (iii) *it is required that $\phi(t; \lambda)$, the characteristic function of $F(x; \lambda)$, be either continuous in λ or real-valued for real t). Let three cumulants with orders j, $j + h$, $j + 2h$ (j, h* positive integers) *exist and be non-zero. If j is odd or both j and h are even, also let $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$. Then the only linear combinations of a finite number of independent variables with distributions in the family, $\sum_{r=1}^{k} c_r X_r$, ($c_r \neq 0$, real), whose distributions are also in the family are those with all $c_r = 1$.*

Proof. According to Theorems 1 and 2 of [8] the characteristic function of a member of the family is of the form $[f(t)]^{\lambda}$, where $f(t)$ is a characteristic function not depending on λ. Let λ_r be the value of the parameter of the distribution of X_r. If $\sum_{r=1}^{k} c_r X_r$ is to have its distribution in the family for some λ, then

$$\prod_{r=1}^{k} [f(c_r t)]^{\lambda_r} = [f(t)]^{\lambda}. \tag{1}$$

Since the cumulants of order through $j + 2h \equiv m$ exist,

$$\log f(t) = \sum_{\nu=1}^{m} \frac{\kappa_\nu}{\nu!} (it)^\nu + o(t^m) \tag{2}$$

Received July 12, 1955; revised April 29, 1958.

in some neighborhood of $t = 0$, where the κ_ν are the cumulants of the distribution corresponding to $f(t)$. Hence

$$\sum_{r=1}^{k} \lambda_r \sum_{\nu=1}^{m} \frac{\kappa_\nu}{\nu!} (ic_r t)^\nu + o(t^m) = \lambda \sum_{\nu=1}^{m} \frac{\kappa_\nu}{\nu!} (it)^\nu + o(t^m),$$

or

$$\sum_{\nu=1}^{m} \frac{\kappa_\nu}{\nu!} (it)^\nu \left(\sum_{r=1}^{k} \lambda_r c_r^\nu - \lambda \right) + o(t^m) = 0. \tag{3}$$

For this to be true as t approaches zero the coefficients of $t, t^2, \cdots, t^m$ must be zero. Since κ_j, κ_{j+h}, and κ_{j+2h} are not zero,

$$\sum_{r=1}^{k} \lambda_r c_r^j = \lambda, \qquad \sum_{r=1}^{k} \lambda_r c_r^{j+h} = \lambda, \qquad \sum_{r=1}^{k} \lambda_r c_r^{j+2h} = \lambda. \tag{4}$$

Multiplying both sides of the second equation by 2 and subtracting them respectively from the sums of the corresponding sides of the first and third equations[1] give

$$\sum_{r=1}^{k} \lambda_r c_r^j (1 - c_r^h)^2 = 0. \tag{5}$$

Since $\lambda_r > 0$ and the c_r are real and not zero, this equation and an even value of j imply that $c_r^h = 1$. If in addition h is odd, then $c_r = 1$. If both j and h are even or if j is odd, then the conditions $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$ imply that $c_r > 0$ as shown below. Hence in these cases also it follows that all c_r are unity.

To show that the conditions $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$ imply that no c_r can be negative, we first note that if all c_r were negative then $\Sigma c_r X_r$ would be negative with probability one and hence could not have its distribution in the family. We therefore suppose that there are exactly p negative values of c_r with $0 < p < k$, say $c_1, c_2, \cdots, c_p$. Let

$$X = -\sum_{r=1}^{p} c_r X_r, \qquad Y = \sum_{r=p+1}^{k} c_r X_r. \tag{6}$$

The cumulative distribution functions of X and Y are, say,

$$\begin{aligned} G(x) &= F\left(-\frac{x}{c_1}; \lambda_1\right) * \cdots * F\left(-\frac{x}{c_p}; \lambda_p\right), \\ H(y) &= F\left(\frac{y}{c_{p+1}}; \lambda_{p+1}\right) * \cdots * F\left(\frac{y}{c_k}; \lambda_k\right), \end{aligned} \tag{7}$$

and thus possess the properties of $F(x; \lambda)$:

$$\begin{aligned} G(x) &= 0 \text{ for } x < 0, \qquad G(x) > 0 \text{ for } x > 0; \\ H(y) &= 0 \text{ for } y < 0, \qquad H(y) > 0 \text{ for } y > 0. \end{aligned} \tag{8}$$

[1] This method of combination, simpler than that used initially, was pointed out by Professor Arne Magnus.

Then

$$(9) \qquad \Pr\left(\sum_{r=1}^{k} c_r X_r < 0\right) = \Pr(Y < X) = \int_0^\infty H(x)\, dG(x).$$

$G(x)$ is not a degenerate distribution since $c_r \neq 0$ and its second cumulant, for example, is not zero. Hence $G(x)$ has a positive increase over some interval in which $H(x) > 0$. Hence there is a positive probability that $\Sigma c_r X_r < 0$. But this is impossible for any member of the family. Hence no c_r can be negative.

3. Discussion of theorem. The requirement that the initial point of increase of the distributions be zero can be dropped by restricting consideration to positive c_r.

The theorem is satisfied with a minimum number of cumulants required if the first three—the mean, the variance, and the "skewness" measure κ_3—are not zero, provided that $F(x, \lambda) = 0$ for $x < 0$ and $F(x, \lambda) > 0$ for $x > 0$. Beyond this proviso, only the requirement $\kappa_3 \neq 0$ need be stated explicitly since this implies $\kappa_2 \neq 0$, and since a non-negative, non-degenerate random variable must have $\kappa_1 \neq 0$. However, the condition $\kappa_3 \neq 0$ is not necessary for the conclusion of the theorem, as shown by the example of the additively closed family of binomial distributions with $p = \frac{1}{2}$ and parameter the sample size. Although $\kappa_3 = 0$ in this case, the theorem applies with κ_2, κ_4, and κ_6 all non-zero.

If the three non-zero cumulants used in the theorem include κ_2, it need not be explicitly stated that $\kappa_2 \neq 0$ since $\kappa_{j+2h} \neq 0$ implies $\kappa_2 \neq 0$.

A requirement that $\kappa_1 \neq 0$ would by itself exclude two cases for which the conclusion of the theorem is false: normal with mean zero and variance λ; Cauchy with median zero and semi-interquartile range λ. However, even $\kappa_1 \neq 0$ and the further conditions $\kappa_2 \neq 0$ and $c_r > 0$ are not sufficient for the conclusion of the theorem. Consider the one-parameter family of normal distributions with variance λ and mean $\gamma\lambda$, where $\lambda > 0$ and $\gamma \neq 0$. The distribution of $c_1X_1 + c_2X_2$ is normal with mean γ times the variance if

$$c_1 = \tfrac{1}{2}[1 + (1 + 4a/\lambda_1)^{\frac{1}{2}}], \qquad c_2 = \tfrac{1}{2}[1 \pm (1 - 4a/\lambda_2)^{\frac{1}{2}}],$$

where $0 < a \leqq \lambda_2/4$. Thus this family does not satisfy the conclusion of the theorem although $\kappa_1 \neq 0$, $\kappa_2 \neq 0$, $c_1 > 0$, $c_2 > 0$.

This example leads to the conclusion that, among distributions with moments of all orders, the condition that some three cumulants of the form κ_j, κ_{j+h}, and κ_{j+2h} not be zero, although it has not been shown necessary for the conclusion of the theorem, is little stronger than necessary. Specifically, we can prove that any additively closed family of non-degenerate distributions with all moments existing (and, in case (iii), characteristic function continuous in λ) which satisfies the conclusion of the theorem must have at least one $\kappa_j \neq 0$ with $j > 2$. For suppose this were not true. Since all moments exist, so do all cumulants. All cumulants beyond κ_2 would be zero. Consequently the distributions would be normal. By Teicher's Theorem 1 [8] a one-parameter additively closed family of non-degenerate normal distributions with, in case (iii), characteristic func-

tions continuous in $\lambda(\lambda > 0)$ must have those characteristic functions of the form

$$(10) \qquad [f(t)]^{\lambda} = e^{\mu it - \sigma^2 t^2/2},$$

where $\sigma > 0$ and $f(t)$ is a characteristic function not depending on λ. Hence

$$(11) \qquad \sigma^2 = \alpha\lambda, \qquad \mu = \gamma\lambda,$$

where $\alpha > 0$ and γ is real. (In case (iii) we may let $\alpha = 1$ without loss of generality.) Equations (10) and (11) would thus be true if the statement in question were not true. But, as shown in the preceding paragraph (where we may take the variance as $\alpha\lambda > 0$ also), this family does not satisfy the conclusion of the theorem, contrary to the hypothesis. Hence the statement in question must be true.

Furthermore, any asymmetrical distribution with characteristic function expansible in a convergent Maclaurin series must have *some* non-zero κ_j for $j > 2$ (or non-zero central moment) of odd order; this follows from the formula for the cumulative distribution function in terms of the characteristic function.

The distributions of some additively closed families are members of the Pearson system. By use of Kendall's recurrence relation for the cumulants of Pearson curves [7] it can be shown that any Pearson-type distribution except the normal for which the recurrence relation is valid has at least three non-zero cumulants of the form κ_j, κ_{j+h}, and κ_{j+2h}. A family of Pearson Type III distributions with left-hand endpoint at zero and the non-additive parameter fixed (the family of all χ^2 distributions for example) is thus an additively closed family that satisfies the theorem.

An example showing that the conditions $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$ are not implicit in the conclusion of the theorem is the family of Poisson distribution functions $F(x; \lambda, b, a)$ where $\lambda > 0$, $b \neq 0$, $a \neq 0$, and F is a step-function with a jump equal to $\lambda^{\nu} e^{-\lambda}/\nu!$ at $x = \lambda b + \nu a$, $\nu = 0, 1, 2, \cdots$ [3]. None of the cumulants is zero, so that the theorem can be applied without invoking the above conditions. Such translation can be applied more generally to additively closed families of distributions; the corresponding slight extension of the theorem is omitted.

It may be questioned whether the existence of any moments is necessary to assure the conclusion of the theorem. In cases (ii) and (iii) the general form of the characteristic function of an infinitely divisible distribution is available [8] and might be thought applicable. No appreciable results have been derived therefrom, however. The above example of Cauchy distributions shows that a restriction of some sort must be placed on an additively closed family whose moments do not exist in order to assure the conclusion.

4. Further examples. The generalized Poisson distributions associated with an arbitrary but fixed distribution [4] form a one-parameter additively closed family of distributions. The generalized Poisson distributions include the negative binomial distributions [1] and Neyman's contagious distributions [4].

The example in section 3 of the additively closed family of normal distributions not satisfying the conclusion of our theorem can be generalized to certain families of stable distributions. Suppose that we have an additively closed family of stable distributions with additive parameter λ of any of the three types in the theorem, $\phi(t; \lambda)$ being continuous in λ for each t if λ is of type (iii). It follows from the general form of the characteristic function of a stable distribution [5] and from Teicher's Theorem 1 that

$$\log (\phi(t; \lambda)) = \beta(\lambda)it - \theta(\lambda) \mid t \mid^{\alpha(\lambda)} \left[1 + i\delta(\lambda) \frac{t}{|t|} \omega(t; \lambda)\right]$$

$$= \lambda\beta(1)it - \lambda\theta(1) \mid t \mid^{\alpha(1)} \left[1 + i\delta(1) \frac{t}{|t|} \omega(t; 1)\right],$$

where $\alpha(\lambda)$, $\beta(\lambda)$, $\delta(\lambda)$, $\theta(\lambda)$ are real functions of λ satisfying $0 < \alpha(\lambda) \leqq 2$, $|\delta(\lambda)| \leqq 1$, and $\theta(\lambda) \geqq 0$, and where $\omega(t; \lambda) = \tan [\pi\alpha(\lambda)/2]$ or $(2/\pi) \log |t|$ according as $\alpha(\lambda) \neq 1$ or $\alpha(\lambda) = 1$.

By equating real and imaginary parts and simple computations, it is readily established that

$$\theta(\lambda) = \lambda\theta(1), \qquad \alpha(\lambda) = \alpha(1),$$

$$\beta(\lambda) = \lambda\beta(1), \qquad \delta(\lambda) = \delta(1).$$

With these conditions the corresponding family of stable distributions is indeed additively closed. Every stable distribution is in at least one of the infinitely many such families of stable distributions. If X_1 and X_2 are independent random variables whose distributions are in the above family, it can be shown that there exist constants c_1 and c_2 unequal to 0 or 1 such that $c_1X_1 + c_2X_2$ is also in the family. Thus one-parameter additively closed families of stable distributions, with $\phi(t; \lambda)$ continuous in λ in case (iii) of the theorem, cannot satisfy the conclusion of the theorem.

5. Acknowledgment. The author is indebted to the referee for many clarifying and stimulating comments, and to M. M. Siddiqui for comments on a revised version.

REFERENCES

[1] E. Cansado Maceda, "On the compound and generalized Poisson distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 414–416.

[2] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," *Proc. Cambridge Phil. Soc.*, Vol. 30 (1934), pp. 178–191, 179 in particular.

[3] H. Cramér, "Problems in probability theory," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 165–193.

[4] W. Feller, "On a general class of 'contagious' distributions," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 389–400.

[5] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent*

Random Variables, Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954, p. 162.
[6] G. S. JAMES, "Notes on a theorem of Cochran," *Proc. Cambridge Phil. Soc.*, Vol. 48 (1952), pp. 443–446.
[7] M. G. KENDALL, "A recurrence relation for the semi-invariants of Pearson curves," *Biometrika*, Vol. 32 (1941), pp. 81–82.
[8] H TEICHER, "On the convolution of distributions," *Ann. Math. Stat.*, Vol 25 (1954), pp. 775–778.

NOTES

NOTE ON RELATIVE EFFICIENCY OF TESTS

By Colin R. Blyth[1]

University of Illinois and Stanford University

1. Summary. This note is concerned with possible definitions of relative efficiency for two sequences of tests of the same hypothesis. For two examples of one kind of definition, relative efficiencies of the Student test and sign test against normal alternatives are calculated for fixed sample size and asymptotically.

2. Introduction. Consider the following problem of relative efficiency of tests. Experiments $X_1, X_2, \cdots$ and two sequences $\{A_n(X_1, \cdots, X_n)\}$, $\{A_n^*(X_1, \cdots, X_n)\}$ of level α tests are available for testing the same hypothesis. We must decide whether to use an A test or an A^* test. Commonly one sequence, say the A^*'s, gives better power for given sample size, but for some reason such as wider validity we may prefer one of the "less efficient" A tests.

The general decision formulation for this problem would use three loss functions (i) cost of experimentation (ii) loss from wrong decisions (iii) disadvantages of using A^*. The usual kinds of decision problems for three loss functions could then be discussed. In practice (iii) is hard to assess and there is no natural comparability between (i) and (ii). So what is usually done is to consider (i) and (ii) only, and having required a bound on one of them, to decide whether the decrease in the other is enough to compensate for the disadvantages of using A^* instead of A. More specifically, the following two types of problems are of interest.

(a). *Fixed power requirement problems.* For a given power requirement, shall we use A_n or $A_{n^*}^*$? Here n and n^* are the smallest sample sizes for which the respective kinds of tests satisfy the given power requirement. Some function $K(n, n^*)$ such as $C(n) - C(n^*)$ or $1 - C(n^*)/C(n)$ is chosen as measuring our loss (extra experimentation cost) from using A_n instead of $A_{n^*}^*$. If $K(n, n^*)$ is small enough we will prefer to use A_n because of the advantages (iii) of A tests. If the given power requirement is a function of an unknown parameter θ, the loss $K(n, n^*)$ will also be a function of θ and so cannot be used directly for deciding between A_n and $A_{n^*}^*$. Some measure of loss not dependent on θ is needed. One natural choice is the worst possible loss $\sup_\theta K(n, n^*)$. (Weighted averages over θ and limits over particular sequences of θ's have also been used.) Asymptotic behavior of $K(n, n^*)$ and $\sup_\theta K(n, n^*)$ can be investigated for sequences of power requirements forcing $n \to \infty$ and $n^* \to \infty$. The particular choice $K(n, n^*) = 1 - n^*/n$ (with n^*/n being called the efficiency of A relative to A^*) and its asymptotic properties has been of wide interest [1], [2], [3], [4].

Received December 9, 1957; revised April 25, 1958.

[1] Work supported by the Office of Naval Research.

(b). *Fixed sample size problems.* For a given sample size n shall we use A_n or A_n^*? Let β_n be the power of A_n and β_n^* the power of A_n^*. Some function $L(\beta_n, \beta_n^*)$ such as $\beta_n^* - \beta_n$ or $1 - \beta_n/\beta_n^*$ is chosen as measuring our loss (extra wrong decisions) from using A_n instead of A_n^*. If $L(\beta_n, \beta_n^*)$ is small enough we will prefer to use A_n because of the advantages (iii) of A tests. If the powers β_n and β_n^* are functions of an unknown parameter θ, the loss $L(\beta_n, \beta_n^*)$ will also be a function of θ and so cannot be used directly for deciding between A_n and A_n^*. Some measure of loss not dependent on θ is needed. One natural choice is the worst possible loss $\sup_\theta L(\beta_n, \beta_n^*)$. This choice appears, with $L = \beta^* - \beta$, in the definition of stringency. Asymptotic behavior of $L(\beta_n, \beta_n^*)$ and $\sup_\theta L(\beta_n, \beta_n^*)$ as $n \to \infty$ can be investigated.

Though interest has been mostly in type (a) problems, it would seem that type (b) problems should be about equal in interest and applicability. The purpose of the present note is to discuss, as an illustration of type (b) problems, the following simple example.

3. Sign Test vs. Student Test. Let $X_1, X_2, \cdots$ be independent, each with Normal (θ, σ^2) distribution. We are to test at level α the one-sided hypothesis $\{\theta \leqq 0\}$ against the alternative $\{\theta > 0\}$. Let $\delta = \theta/\sigma$ and $p = p(\delta) = P(X_i > 0) = F(\delta)$ where F is the Normal (0, 1) cumulative. Then the number R_n of positive observations among $X_1, \cdots, X_n$ has a Binomial (n, p) distribution. And

$$T_n = n^{\frac{1}{2}}\bar{X}/[\sum(X_i - \bar{X})^2/(n-1)]^{\frac{1}{2}}$$

has a Student t distribution with $n - 1$ degrees of freedom which is central when $\delta = 0$ and non-central with parameter $n^{\frac{1}{2}}\delta$ in general.

The sign test A_n of $\{\theta \leqq 0\}$ is

$$\begin{cases} \text{Reject when } R_n - n/2 > k_n \\ \text{Reject with prob. } \gamma_n \text{ when } R_n - n/2 = k_n, \end{cases}$$

where k_n, γ_n are constants determined by

$$P(R_n - n/2 > k_n \mid \delta = 0) + \gamma_n P(R_n - n/2 = k_n \mid \delta = 0) = \alpha.$$

The power function of this test is

$$\beta_n(\delta) = P(R_n - n/2 > k_n) + \gamma_n P(R_n - n/2 = k_n).$$

Values of $k_n, \gamma_n, \beta_n(\delta)$ can be obtained from tables such as [5] of the binomial distribution. For large values of n the normal approximation to binomial gives

$$\beta_n(\delta) \cong F\left(\frac{\sqrt{n}\,(2p-1) - c}{2\sqrt{p(1-p)}}\right) \quad \text{where} \quad F(c) = 1 - \alpha. \tag{1}$$

The Student test A_n^* of $\{\theta \leqq 0\}$ is

$$\text{Reject when } T_n > c_n$$

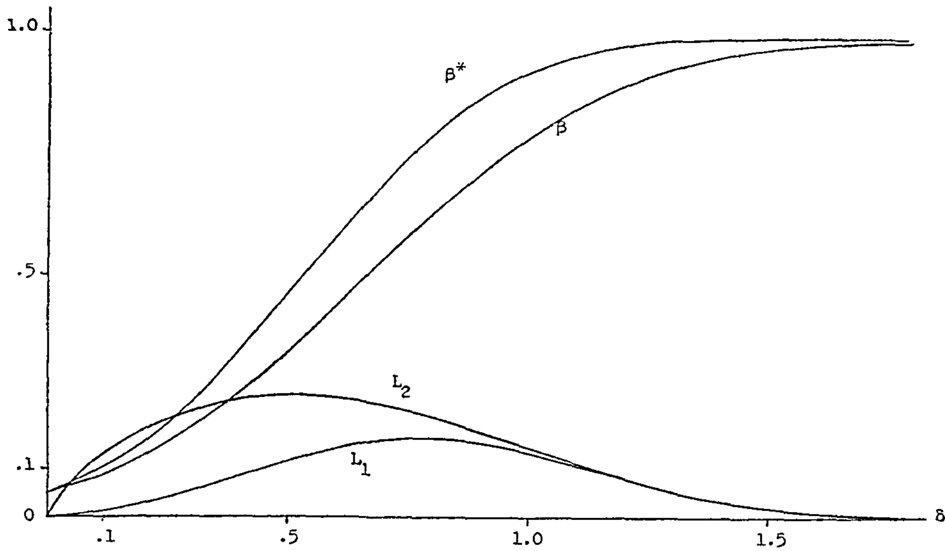


FIG. 1. Power functions β^* of Student test and β of sign test for $\alpha = .05$, $n = 11$; $L_1 = \beta^* - \beta$, $L_2 = 1 - \beta/\beta^*$.

where c_n is a constant determined by

$$P(T_n > c_n \mid \delta = 0) = \alpha.$$

The power function of this test is

$$\beta_n^*(\delta) = P(T_n > c_n) .$$

Values of c_n can be obtained from tables of the Student t distribution, and values of $\beta_n^*(\delta)$ from tables such as [6] of the non-central Student t distribution. For large values of n the normal approximation to non-central Student t gives

$$\beta_n^*(\delta) \cong F(\sqrt{n}\delta - c) \quad \text{where} \quad F(c) = 1 - \alpha. \tag{2}$$

Loss functions such as

$$L_1^n(\delta) = L_1(\beta_n, \beta_n^*) = \beta_n^*(\delta) - \beta_n(\delta)$$

$$L_2^n(\delta) = L_2(\beta_n, \beta_n^*) = 1 - \beta_n(\delta)/\beta_n^*(\delta)$$

can easily be plotted for particular values of n and α. This is done in Figure 1 for $n = 11$, $\alpha = .05$. As δ increases from 0 each function $L_i(\delta)$, $i = 1, 2$ increases from 0 to a maximum and then decreases toward 0.

For fixed α the change in appearance of these curves with increasing n differs only slightly from a simple horizontal compression. The curve $L_i^n(\delta)$ rises more quickly to its maximum and then falls more quickly toward 0, with increasing n. The position of the maximum tends to 0 at the rate $1/n^{\frac{1}{2}}$ but the maximum value changes very little and has a limit. Table 1 gives values of $\sup_\delta L_i^n(\delta)$ for $\alpha = .05$ and $n = 2, 3, \cdots, 13$. These values are computed from tables [5], [6] using interpolation and should be in error by not more than one or two units in the third decimal place. The cases $n = 2, 3, 4$ are special because for these the sign test does not reject with probability 1 even when $R_n = 0$ and so the

TABLE 1

Maxima of L_1 = power loss, L_2 = 1 − power ratio, of sign test relative to Student test, $\alpha = .05$

n	sup L_1	sup L_2
2	.800	.800
3	.600	.600
4	.200	.200
5	.130	.197
6	.189	.263
7	.150	.212
8	.153	238
9	.180	.261
10	.142	.213
11	.167	.252
12	171	.260
13	.151	227
∞	.1686	.2610

power of the sign test does not $\to 1$ as $\sigma \to \infty$. For $n = 5, 6, \cdots \sup_\delta L_i^n(\delta)$ tends to be smaller if there is a non-randomized sign test with size close to.05 [$n = 5, 8, 10, 13$] and larger if there is no such sign test [$n = 6, 9, 12$]. Even for the smallest of these n the differences from the asymptotic values $\lim_{n\to\infty} \sup_\delta L_i^n(\delta)$ are not large.

Discussion of this example is concluded with the calculation of these asymptotic values. The following easily proved result is used:

LEMMA.

$$\lim_{n\to\infty} \sup_\delta L_n(\delta) = \sup_\Delta \lim_{n\to\infty} L_n(\delta_n)$$

if the former exists, where Δ is the set of all sequences $\{\delta_n\}$ for which $\lim_{n\to\infty} L_n(\delta_n)$ *exists.* [If lim be replaced throughout by lim inf or lim sup the same result holds, with existence provisos unnecessary.]

Writing $\delta_n = a_n/n^{\frac{1}{2}}$ it easily follows from (1) and (2) that if $a_n \to a$ then

$$\beta_n(\delta_n) \to F(a\sqrt{2/\pi} - c), \qquad \beta_n^*(\delta_n) \to F(a - c)$$

where $F(c) = 1 - \alpha$. This gives

$$\lim_{n\to\infty} L_1^n(\delta_n) = F(a - c) - F(a\sqrt{2/\pi} - c) \tag{3}$$

$$\lim_{n\to\infty} L_2^n(\delta_n) = 1 - F(a\sqrt{2/\pi} - c)/F(a - c). \tag{4}$$

Because of the lemma we can find $\lim_{n\to\infty} \sup_\delta L_i^n(\delta), i = 1, 2$ by finding the value of a giving a maximum in (3), (4). Differentiating (3) with respect to a and equating the result to zero gives

TABLE 2

Asymptotic maximum power loss R_1 and proportionate power loss R_2 for sign test relative to Student test

α	a'	a''	R_1	R_2
.25	1.5514	1.1784	.0963	.1268
.10	1.6245	1.4086	.1405	.2056
.05	2.3570	1.5593	.1686	.2610
.01	3.0019	1.8574	.2229	.3765
.001	3.7676	2.2087	.2844	.5128
0	∞	∞	1	1

$$(3') \qquad \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(a-c)^2\} = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(a\sqrt{2/\pi}-c)^2\}$$

which reduces to

$$(a-c)^2 = (a\sqrt{2/\pi}-c)^2 + \log(\pi/2).$$

The root of this quadratic at which the maximum of (3) occurs is

$$a' = \frac{c}{1+\sqrt{2/\pi}}\{1+\sqrt{1+(\log \pi/2)(1+\sqrt{2/\pi})/(1-\sqrt{2/\pi})c^2}\}.$$

The maximum value $R_1 = \lim_{n\to\infty} \sup_\delta L_1^n(\sigma)$ can now be found by substituting a' for a in (3). For example $\alpha = .05$ gives $c = 1.6449$, $a' = 2.3750$, and $R_1 = .1686$ for the asymptotic maximum loss. Differentiating (4) with respect to a and equating the result to zero gives

$$(4') \qquad F(a-c)\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{2\pi}}\exp\{-\tfrac{1}{2}(a\sqrt{2/\pi}-c)^2\} = F(a\sqrt{2/\pi}-c)\frac{1}{\sqrt{2\pi}}\exp\{-\tfrac{1}{2}(a-c)^2\}.$$

For given α the solution a'' of (4′) can be found numerically and shown to maximize (4). The maximum value $R_2 = \lim_{n\to\infty} \sup_\delta L_2^n(\delta)$ can now be found by substituting a'' for a in (4). For example $\alpha = .05$ gives $c = 1.6449$, $a'' = 1.5593$, and $R_2 = .2610$ for the asymptotic maximum loss.

Table 2 gives a', a'', R_1 (the asymptotic maximum power loss), R_2 (the asymptotic maximum amount by which the power ratio falls below 1) for several values of α. The most noticeable feature of this table is the strong dependence of R_1 and R_2 on the value of α. For small α use of sign test instead of Student test results in very severe loss of power at some alternatives. For example when $\alpha = .001$ there is an alternative where 51% of the power is lost by using sign test instead of Student test, and an alternative where the amount of power lost is .28.

REFERENCES

[1] E. J. G. Pitman, *Lecture notes on nonparametric statistical inference*, Columbia University, 1949.

[2] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat*, Vol. 23 (1952), pp. 493–507.

[3] W. J. Dixon, "Power functions of the sign test and power efficiency for normal alternatives," *Ann. Math. Stat*, Vol 24 (1953), pp 467–473.

[4] J. L. Hodges, Jr. and E. L. Lehmann, "The efficiency of some nonparametric competitors of the t-test," *Ann. Math. Stat*, Vol 27 (1956), pp. 324–335.

[5] Harvard Computation Laboratory, *Tables of the cumulative binomial probability distribution*, Harvard University Press, 1955

[6] J. Neyman and B. Tokarska, "Errors of the second kind in testing Student's hypothesis," *Jour. Am. Stat. Assoc*, Vol 31 (1936), pp 318–326.

A NOTE ON CONFIDENCE INTERVALS IN REGRESSION PROBLEMS

By John Mandel

National Bureau of Standards

This note deals with the construction of confidence intervals for arbitrary real functions of multiple regression coefficients

Consider the usual model

$$y_\alpha = \sum_i \beta_i x_{i\alpha} + \epsilon_\alpha \qquad \begin{array}{l} i = 1, \cdots, k \\ \alpha = 1, \cdots, N \end{array} \tag{1}$$

in which the ϵ_α are independently and normally distributed with mean zero, and common variance σ^2.

It is customary to construct confidence intervals for the β_i, using Student's t distribution. Alternatively, a joint confidence region can be constructed for the β_i using critical values of the F distribution. In both cases the usual statistic s^2, based on $N - k$ degrees of freedom, is used as an estimate of σ^2.

Durand [1] has discussed the use of the joint confidence region of the β_i, an ellipsoid in a k-dimensional space, for the construction of confidence intervals for linear functions, $Q = \sum_i h_i\beta_i$ of the regression coefficients. He points out that the chosen confidence coefficient (corresponding to the ellipsoid) is a lower bound for the joint confidence of any set of intervals thus derived.

Our first objective is to generalize this procedure by removing the restriction of linearity. Let

$$z = f(\beta_1, \beta_2, \cdots, \beta_k) \tag{2}$$

be any real function of the coefficients β_i. The form of the function is arbitrary but known.

For any arbitrarily selected value of z, say z_0, equation (2) represents a

Received February 15, 1957; revised March 1, 1958.

hypersurface in the k-dimensional parameter space of the β_i. Denote by $M[z]$ the set of all values of z_0 for which the corresponding hypersurfaces "cut" the ellipsoid, i.e., for which the equation:

$$z_0 = f(\beta_1, \beta_2, \cdots, \beta_k)$$

and the quadratic equation representing the ellipsoid have at least one common real solution in the β_i.

The set $M[z]$ is, in general, a closed interval, bounded by those two values of z for which the corresponding hypersurfaces are tangent to the ellipsoid. Furthermore, the event that the point corrresponding to the "true" values of the β_i is inside the ellipsoid implies that the z-value corresponding to these true values is an element of $M[z]$, but the converse is not necessarily true. Consequently, since the probability of the former event is equal to the confidence coefficient $1 - \alpha$, the probability of the latter event is at least $1 - \alpha$. If other functions $u = \varphi(\beta_1, \beta_2, \cdots, \beta_k)$, $v = \psi(\beta_1, \beta_2, \cdots, \beta_k)$, etc., are considered simultaneously with z, it follows that the confidence intervals constructed by the above procedure for $z, u, v, \cdots$ are all jointly valid with a joint confidence for which $1 - \alpha$ is a lower bound.

Our next objective is to discuss, in the light of the above procedure, a regression problem often encountered in practice.

Consider the straight line regression

$$y_\alpha = \beta_0 + \beta_1(x_\alpha - \bar{x}) + \epsilon_\alpha \qquad \alpha = 1, 2, \cdots, N \tag{3}$$

where $\bar{x} = (1/N)\sum_\alpha x_\alpha$. Having obtained least squares estimates for β_0 and β_1, say b_0 and b_1, consider p "future" observations of y and let it be required to find confidence intervals for the corresponding p values of x.

This problem involves, in addition to the random errors of the original N values of y, as reflected in the random fluctuations of the least squares estimates b_0 and b_1, also the random errors of the p "future" y values. Denote the "future" observations by $y_{N+1}, y_{N+2}, \cdots, y_{N+p}$, and their expected values by $\eta_{N+1}, \eta_{N+2}, \cdots, \eta_{N+p}$. Consider the $p + 2$ dimensional space with coordinates β_0, $\beta_1, \eta_{N+1}, \eta_{N+2}, \cdots, \eta_{N+p}$. The joint confidence ellipsoid for these $p + 2$ values, for any given confidence coefficient, will be centered on $b_0, b_1, y_{N+1}, y_{N+2}, y_{N+p}$, and can be found as follows by a generalization of a method used by Working and Hotelling [8]:

The quantity

$$\chi_1^2 = \frac{(\beta_0 - b_0)^2}{\sigma_{b_0}^2} + \frac{(\beta_1 - b_1)^2}{\sigma_{b_1}^2} + \frac{\sum_{i=1}^{p}(\eta_{N+i} - y_{N+i})^2}{\sigma^2} \tag{4}$$

has the chi-square distribution with $p + 2$ degrees of freedom. $\sigma_{b_0}^2$ and $\sigma_{b_1}^2$ are of course known functions of σ^2, $\sigma_{b_0}^2 = \sigma^2/N$ and $\sigma_{b_1}^2 = \sigma^2/\sum_{i=1}^{N}(x_i - \bar{x})^2$.

On the other hand, we have

$$\chi_2^2 = \frac{(N - 2)s^2}{\sigma^2} \tag{5}$$

a quantity distributed as chi-square with $(N - 2)$ degrees of freedom.

Since χ_1^2 and χ_2^2 are mutually independent, it follows from (4) and (5) that a joint confidence region, with coefficient $1 - \alpha$ for β_0, β_1, and the expected values of $y_{N+1}, \cdots, y_{N+p}$ is given by

$$(6) \qquad \frac{(\beta_0 - b_0)^2}{1/N} + \frac{(\beta_1 - b_1)^2}{1/\sum(x - \bar{x})^2} + \sum_{i=1}^{p} (\eta_{N+i} - y_{N+i})^2 = (p + 2)F_\alpha s^2$$

where F_α is the critical value of the F distribution with $p + 2$ and $N - 2$ degrees of freedom, at the α level of significance.

Consider now the function

$$x' = \bar{x} + \frac{\eta' - \beta_0}{\beta_1}$$

where η' is the expected value of one of the p "future" observations, and x' the corresponding true x-value. By the method previously outlined, confidence limits for x' are obtained by determining the two values of x' for which the hyperplane

$$(7) \qquad \eta' - \beta_0 = \beta_1(x' - \bar{x})$$

is tangent to the ellipsoid, provided that the set of values of x' for which the hyperplane (7) intersects the ellipsoid is a closed interval.

Denoting these limits by x'_L and x'_U, it is found that the quantities $u_L = x'_L - \bar{x}$ and $u_U = x'_U - \bar{x}$ are the roots of the equation

$$(8) \quad \left(b_1^2 - \frac{K^2}{\sum(x - \bar{x})^2}\right) u^2 - 2b_1(y' - b_0)u + \left[(y' - b_0)^2 - \frac{N + 1}{N} K^2\right] = 0$$

where $K^2 = (p + 2)F_\alpha s^2$.

The condition for equation (8) to have distinct real roots is

$$(9) \qquad \frac{(y' - b_0)^2}{\sum(x - \bar{x})^2} + \frac{N + 1}{N}\left[b_1^2 - \frac{K^2}{\sum(x - \bar{x})^2}\right] > 0$$

Condition (9) is necessary but not sufficient for obtaining a confidence interval for x'. This is apparent from the fact that when x' is made $\pm\infty$, equation (7) represents the hyperplane $\beta_1 = 0$. Consequently, if the hyperplane $\beta_1 = 0$ intersects the ellipsoid, the parameter x' will have a discontinuity when (7) becomes $\beta_1 = 0$, and the roots x'_L and x'_U, though distinct and real, will then not be the limits of a confidence interval for x'.

The condition for $\beta_1 = 0$ not to intersect the ellipsoid is

$$(10) \qquad b_1^2 \sum(x - \bar{x})^2 > K^2$$

It can be proved that condition (10), which implies (9), is both necessary and sufficient in order that the roots of (8) yield the limits of a confidence interval for x'.

If equation (10) is satisfied, the procedure leading to equation (8) can also be carried out for the remaining $p - 1$ "future" measurements, y'', y''', etc. In this manner one will obtain a set of confidence intervals (x'_L, x'_U), (x''_L, x''_U),

(x_L''', x_U'''), etc., all of which are jointly valid with a confidence coefficient for which $1-\alpha$ is a lower bound. Furthermore, this lower bound will still apply if confidence intervals are also derived for any number of real functions of β_0, β_1 and the p values $\eta_{N+1}, \eta_{N+2}, \cdots, \eta_{N+p}$.

Equation (8) should be compared to the relation obtained by the use of Fieller's theorem [3, 4]. This theorem leads to a confidence interval for $x' - \bar{x}$ by considering it as the ratio of the two normally distributed variables $y' - b_0$ and b_1, whose variances are $(N + 1)\sigma^2/N$ and $\sigma^2/\sum(x - \bar{x})^2$ and whose covariance is zero. The confidence interval, with coefficient $1 - \alpha$, thus found is given by the roots of the equation

$$(11) \quad \left(b_1^2 - \frac{t_\alpha^2 s^2}{\sum(x - \bar{x})^2}\right)u^2 - 2b_1(y' - b_0)u + \left[(y' - b_0)^2 - \frac{N+1}{N} t_\alpha^2 s^2\right] = 0$$

where t_α is the critical value of Student's t, at the two-sided α level, and u is defined as above.

The only difference between equations (8) and (11) is the substitution of K^2 for t^2s^2, i.e., the substitution of $[(p + 2)F_\alpha]^{\frac{1}{2}}$ for t_α. This substitution results in a widening of the confidence interval, caused by the joint consideration of $p + 2$ parameters instead of the single parameter η', (or its corresponding x'). It is of interest to observe that the relation between $[(p + 2)F_\alpha]^{\frac{1}{2}}$ and t_α is precisely that found by Scheffé [6] in establishing simultaneous confidence statements for all means in an analysis of variance, as contrasted with individual confidence statements based on Student's t.

In deciding whether in a particular application, joint or single confidence intervals should be used, one may be guided by the following plausible rule. Joint confidence intervals are indicated in situations involving two or more quantities that are determined as so many phases of a single problem. On the other hand, quantities involved in unrelated problems, even though they are derived from the same basic data, should not be treated jointly in deriving confidence intervals. It appears advisable, in view of this rule, to partition all the quantities derived from a single set of data into groups such that the quantities within a group—inasmuch as they correspond to the same problem, are treated jointly for the derivation of confidence intervals; while the groups themselves are treated independently of each other.

Groups involving single predictions should be treated by Fieller's theorem, since there appears to be no justification, in such cases, for widening the confidence interval through inclusion of confidence statements about the slope and the intercept.

It is of interest to note that the confidence interval based on equation (8) may be obtained by drawing hyperbolic confidence limits [2] for the straight line represented by equation (3), in accordance with the relations

$$(12) \qquad y = b_0 + b_1(x - \bar{x}) \pm K\left[\frac{N+1}{N} + \frac{(x - \bar{x})^2}{\sum(x - \bar{x})^2}\right]^{1/2}$$

and by determining the x-interval defined by the intersection of the line $y = y'$

with the two branches of this hyperbola. It is readily seen that the condition that such an x-interval exists and be of finite length is equivalent to the condition that the two asymptotes of the hyperbola have slopes of equal sign. Since these slopes are $b_1 - K/[\sum(x - \bar{x})^2]^{\frac{1}{2}}$ and $b_1 + K/[\sum(x - \bar{x})^2]^{\frac{1}{2}}$, the condition in question is $b_1^2 - K^2/(\sum(x - \bar{x})^2) > 0$. This is condition (10) obtained previously by a different line of reasoning.

It may be observed, finally, that the inverse problem, viz, to determine uncertainty intervals for observed y values corresponding to given x values [2] is not a classical case of interval estimation, since it is concerned with bracketing a random variable, not a population parameter, by means of two statistics. Intervals of this type are discussed by Weiss [7].

Applications of the procedure outlined in this note to a problem in chemistry are discussed elsewhere [5].

REFERENCES

[1] D. Durand, "Joint Confidence Regions for Multiple Regression Coefficients," *J Am Stat. Assoc.*, 49, 130 (1954)

[2] C. Eisenhart, "The Interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research," *Ann. Math Stat*, 10, 162 (1939).

[3] E. C. Fieller, "Biological Standardization of Insulin," *J Roy. Stat. Soc*, Supplement 7, 1 (1940).

[4] D. J. Finney, *Probit Analysis*, Cambridge University Press, Cambridge, 1952.

[5] J. Mandel and F. J. Linnig, "Study of Accuracy in Chemical Analysis Using Linear Calibration Curves," Analytical Chemistry, 29, 743 (1957).

[6] H. Scheffé, "A Method for Judging all Contrasts in the Analysis of Variance," *Biometrika*, 40, 87 (1953).

[7] L. Weiss, "A Note on Confidence Sets for Random Variables," *Ann Math Stat*, 26, 142 (1955).

[8] H. Working and H. Hotelling, "Applications of the Theory of Error to the Interpretation of Trends," *J Am Stat Assoc* 165A (Proceedings), 73 (1929).

A NOTE ON INCOMPLETE BLOCK DESIGNS

By A. M. Kshirsagar

University of Bombay

1. Introduction. Kempthorne [1] has shown the efficiency factor of an incomplete block design to be a quantity proportional to the harmonic mean of the non-zero latent roots of the matrix of coefficients of the reduced normal equations for the intra-block estimates of treatment effects. He has further stated that the geometric mean in a certain sense corresponds to the generalized variance but has not explicitly explained it. The present note is intended to clear this point and to prove that the design with highest efficiency factor (in any case, whether the harmonic mean or the geometric mean is taken as a measure of efficiency) is

(a) a balanced incomplete block design, if such a design exists; and

Received April 19, 1957; revised November 15, 1957.

(b) a Youden Square, if it exists, among designs in which heterogeneity is eliminated in two directions.

There is some overlap between this paper and the ones by Kiefer and by Mote in this issue.

2. Incomplete block design. Let there be v treatments and b blocks of k plots each. Let r be the number of replications of each treatment and let N be the incidence matrix of the design (rows refer to the treatments and columns to blocks). Each element of N, for an incomplete block design, is either 0 or 1, Then the matrix of coefficients of the reduced normal equations for the intra-block estimates t_i of the treatment effects is

$$C = rI_v - \frac{1}{k} NN',$$

where I_v denotes the identity matrix of order v. For any design, C has one zero latent root, the corresponding latent vector having all the elements equal. Let the non-zero roots of C be $\lambda_1, \lambda_2, \cdots, \lambda_{v-1}$ and let $m_1, m_2, \cdots, m_{v-1}$ be the corresponding orthogonal normalized latent vectors (column). Kempthorne [1] chooses the average variance of elementary treatment contrasts like $t_i - t_j$ to arrive at the harmonic mean of the λ's as a definition of efficiency factor of a design. The author, however, feels that, instead, a complete set of $v - 1$ orthogonal normalized treatment contrasts be chosen because,

(1) their average variance leads to the harmonic mean of the λ's; and

(2) their generalized variance leads to the geometric mean of the λ's, as a criterion to measure the efficiency of a design. Let $l_i\,(i = 1, 2, \cdots, v-1)$ be orthogonal normalized column vectors so that $l_i't$ where

$$t = \begin{bmatrix} t_1 \\ t_2 \\ \cdots \\ t_v \end{bmatrix}$$

form a complete set of orthogonal normalized treatment contrasts. Then, if we observe that

$$\sum_{i=1}^{v-1} (l_i' t)^2 = \sum_{i=1}^{v} (t_i - \bar{t})^2,$$

$$= \frac{1}{2v} \sum_{\substack{i,j=1 \\ i \neq j}}^{v} (t_i - t_j)^2$$

and use Kempthorne's [1] result about average variance of $t_i - t_j$, it follows readily that the average variance of a full set of orthogonal normalized treatment contrasts is proportional to the hormonic mean of $\lambda_1, \lambda_2, \cdots, \lambda_{v-1}$.

However, if we consider the generalized variance of $l_i't\ (i = 1, 2, \cdots, v-1)$, it can be shown that it is proportional to $(\lambda_1\lambda_2 \cdots \lambda_{v-1})^{-1}$. This can be proved by using the fact that the transformation from $l_i't\ (i = 1, 2, \cdots, v-1)$

to $m_i't$ $(i = 1, 2, \cdots, v-1)$ is orthogonal and that

$$V(m_i' t) = \frac{\sigma^2}{\lambda_i}; \qquad i = 1, 2, \cdots, v-1$$

and

$$\text{Cov}(m_i't, m_j't) = 0, \qquad i \neq j,$$

where σ^2 is the variance of the yield of a plot.

Thus, either $\sum_1^{v-1} 1/\lambda_i$ or $(\lambda_1\lambda_2 \cdots \lambda_{v-1})^{-1}$ can be taken as a measure of efficiency of the design. It should be noted that

$$\sum_{i=1}^{v-1} \lambda_i = \text{trace } C = vr\left(1 - \frac{1}{k}\right).$$

Hence to obtain a design with highest efficiency we have to minimise either $\sum_{i=1}^{v-1} 1/\lambda_i$ or $(\lambda_1\lambda_2 \cdots \lambda_{v-1})^{-1}$ subject to the condition that

$$\sum_{i=1}^{v-1} \lambda_i = \text{constant}.$$

This immediately leads to

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{v-1} = \frac{vr}{v-1}\left(1 - \frac{1}{k}\right)$$

and consequently,

$$C = \frac{vr}{v-1}\left(1 - \frac{1}{k}\right)\left(I_v - \frac{1}{v} E_{vv}\right)$$

where E_{pq} denotes a $p \times q$ matrix, all the elements of which are unity. This proves, therefore, that the design with the highest efficiency is a balanced incomplete block design, if such a design exists.

3. Designs in which heterogeneity is eliminated in two directions. Let there be UU' plots arranged in U rows and U' columns, and let v treatments be assigned to these plots in such a way that every treatment is replicated r times and the ith treatment occurs l_{ij} times in the jth row and m_{ik} times in the kth column $(i = 1, 2, \cdots, v; j = 1, 2, \cdots, U; k = 1, 2, \cdots, U')$ where l_{ij} and m_{ik} are either 0 or 1. Let $L = [l_{ij}]$ and $M = [m_{ik}]$. Then the matrix of coefficients of reduced normal equations for treatments effects after eliminating row and column effects is

$$C_0 = rI_v - \frac{1}{U'} LL' - \frac{1}{U} MM' + \frac{r^2}{UU'} E_{vv}.$$

This matrix C_0 plays the same role as C in section 2. Hence for a design of this type, the efficiency is maximum if all the non-zero latent roots of C_0 are equal,

the common value being

$$\frac{1}{v-1}\text{ trace } C_0 = \frac{vr}{v-1}\left(1 - \frac{1}{U} - \frac{1}{U'} + \frac{1}{UU'}\right),$$
$$= a, \text{ say.}$$

It therefore follows that for designs in which heterogeneity is eliminated in two directions, the efficiency factor is maximum if

$$\frac{1}{U'} LL' + \frac{1}{U} MM' \text{ is of the form}$$

$$\begin{bmatrix} p & q & q & \cdots & q \\ q & p & q & \cdots & q \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ q & q & q & \cdots & p \end{bmatrix}.$$

It should be observed that, for a Youden Square (where the rows are complete blocks and columns form a symmetrical balanced incomplete block design),

$$U = r, \qquad U' = v$$

and

$$L = E_{vv}$$

and

$$MM' = \begin{bmatrix} r & \lambda & \lambda & \cdots & \lambda \\ \lambda & r & \lambda & \cdots & \lambda \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \lambda & \lambda & \lambda & \cdots & r \end{bmatrix}.$$

and $LL'/U' + MM'/U$ is of the required form. Consequently, among designs in which heterogeneity is eliminated in two directions, a Youden Square, if it exists, has maximum efficiency.

Acknowledgement: I am indebted to Prof. M. C. Chakrabarti and the referee for their valuable help and suggestions in the preparation of this note.

REFERENCE

[1] O. Kempthorne, "The efficiency factor of an incomplete block design." *Ann. Math. Stat.* Vol. 27 (1956), pp. 846–849.

ON A MINIMAX PROPERTY OF A BALANCED INCOMPLETE BLOCK DESIGN

By V. L. Mote

Institute of Statistics, Raleigh, North Carolina

Summary. It is shown that for a given set of parameters (b blocks, k plots per block and v treatments), among the class of connected incomplete block designs,

Received July 2, 1957; revised January 15, 1958.

a balanced incomplete block design (if it exists) is the design which maximizes the minimum efficiency, efficiency being defined as

$$\frac{\text{Variance of an estimated treatment contrast in a randomized block}}{\text{Variance of the estimated treatment contrast in the incomplete block}}.$$

The proof will be preceded by a lemma.

Notation. Capital letters will be used to denote matrices and boldface small letters to denote vectors. At times a matrix of m rows and n columns will be denoted by $A(m \times n)$.

LEMMA. *If $B(p \times p)$ is real symmetric and at least positive semidefinite of rank $r(\leqq p)$, then:*

(i) *The stationary values of*

$$\frac{\mathbf{a}'(1 \times p)B(p \times p)\mathbf{a}(p \times 1)}{\mathbf{a}'\mathbf{a}}$$

under the variation of $\mathbf{a}$ *(over all non-null* $\mathbf{a}$ *excepting the solutions of* $B\mathbf{a} = 0$*) are the characteristic roots of B.*

(ii) *In particular the largest and the smallest values of* $\mathbf{a}'B\mathbf{a}/\mathbf{a}'\mathbf{a}$ *(under the variation of all non-null* $\mathbf{a}$ *excepting the solutions of* $B\mathbf{a} = 0$*), are the largest and the smallest non-zero characteristic roots of B.*

(iii) $\mathbf{a}'B\mathbf{a}/\mathbf{a}'\mathbf{a}$ *attains its maximum (or minimum) value if and only if* $\mathbf{a}$ *is a latent vector corresponding to the maximum (or minimum) latent roots of B.*

For a proof of this lemma we refer to S. N. Roy [3] and H. W. Turnbull and A. C. Aitken [4].

Let us adopt the following notation:

$\lambda_{i\alpha}$ = number of blocks in which the ith and the αth treatments appear together.

r_i = number of blocks in which the ith treatment appears.

$$c_{i\alpha} = \begin{cases} \dfrac{-\lambda i\alpha}{k}\, i \neq \alpha;\, i = 1, 2, \cdots\ v;\, \alpha = 1, 2, \cdots, v. \\ r_i\left(1 - \dfrac{1}{k}\right), i = \alpha. \end{cases}$$

T_i = total yield of the ith treatment.

B_j = total yield of the jth block.

$$n_{ij} = \begin{cases} 1 \text{ if the } i\text{th treatment appears in the } j\text{th block,} \\ 0 \text{ otherwise.} \end{cases}$$

$$Q_i = T_i - \frac{1}{k}\sum_{j=1}^{b} n_{ij} B_j.$$

Finally let

$$\mathbf{Q}'(1 \times v) = (Q_1 Q_2 \cdots Q_v).$$

In any connected incomplete block design the adjusted normal equations

are given by

$$Ct = Q$$

where

$$C = (c_{i\alpha}) \qquad i = 1, 2, \cdots, v, \quad \alpha = 1, 2, \cdots, v.$$

It is well known that C is symmetric positive semidefinite of rank $v - 1$ and that the only independent non-trivial solution of the equations $Cx = 0$ is

$$x'(1 \times v) = (1, 1, \cdots, 1).$$

Let $m'(1 \times v) = (m_1 m_2, \cdots, m_v)$ be a non-null vector such that $\sum_{i=1}^{v} m_i = 0$.

It is well known (e.g., see R. C. Bose and S. Ehrenfeld) that the variance of the "best estimate" of $m't$ is given by $\varrho' C \varrho \sigma^2$ where ϱ is a solution of $C\theta = m$.

We shall now show that

$$\sup_{m \epsilon M} \frac{\varrho' C \varrho}{m'm} = \frac{1}{\lambda_{\min}}$$

where M is the class of all non-null vectors $m'(1 \times v) = (m_1, m_2, \cdots, m_v)$ such that $\sum_i m_i = 0$ and $\lambda_{\min}$ is the smallest of the $v - 1$ non-zero characteristic roots of C.

Since C is real symmetric, it follows that there exists an orthogonal matrix $P(v \times v)$ such that

$$P'CP = \begin{bmatrix} D_{\lambda_i}\ [(v-1) \times (v-1)] & 0\ [(v-1) \times 1] \\ 0\ [1 \times (v-1)] & 0 \end{bmatrix}$$

where D_{λ_i} is a diagonal matrix; the diagonal elements being $\lambda_1, \lambda_2, \cdots, \lambda_{v-1}$ the non-zero latent roots of C. Let

$$P = [P_1[v \times (v-1)] \qquad q\ (v \times 1)].$$

Then $C = P_1 D_{\lambda_i} P_1'$.

It can be easily shown that

$$P_1 P_1' + qq' = I, \tag{2}$$

$$P_1' P_1 = I, \tag{3}$$

and that the rank of P_1 is $v - 1$ and

$$q'(1 \times v) = \frac{1}{\sqrt{v}}(1, 1, \cdots, 1). \tag{4}$$

It can be seen that

$$\varrho = [P_1 D_{\lambda_i}^{-1} P_1']m$$

is a solution of $C\theta = m$, and

$$\frac{\varrho' C \varrho}{m'm} = \frac{m' P_1 D_{\lambda_i}^{-1} P_1' m}{m'm}.$$

Hence by virtue of the lemma stated earlier we have

$$\operatorname*{Sup}_{m \epsilon M} \frac{m'(P_1 D_{\lambda_i}^{-1} P_1')m}{m'm} = \frac{1}{\lambda_{\min}}.$$

The variance of the "best estimate" of m′t in a randomized block is

$$(1/b)m'm\sigma^2.$$

Hence,

$$\text{efficiency} = \left(\frac{1}{b}\right)\frac{m'm}{\varrho' C \varrho}$$

where ϱ is a solution of $C\theta = m$. Now

$$\inf_{m \epsilon M}\left[\frac{m'm}{\varrho' C \varrho}\right] = \left[\frac{1}{\operatorname*{Sup}_{m\epsilon M} \frac{\varrho' C\varrho}{m'm}}\right] = \lambda_{\min}.$$

Hence, minimum efficiency $= \lambda_{\min}/b$. It can be shown that for any connected design $\lambda_{\min} \leqq \lambda v/k$, where

$$\lambda = \frac{bk(k-1)}{v(v-1)}.$$

Now if we can show that, $\lambda_{\min} = \lambda v/k$ if and only if the design is a balanced incomplete block design, then our problem is solved. If the design is a balanced incomplete block design, then, $\lambda_{\min} = \lambda v/k$, since $\lambda v/k$ is a latent root of multiplicity $v - 1$ for the C corresponding to the given design. The next thing we have to show is that if $\lambda_{\min} = \lambda v/k$, then the design is a balanced incomplete block design. Since $\lambda_{\min} = \lambda v/k$, it follows that all of the remaining $v - 2$ roots must be exactly $\lambda v/k$. Hence

$$C = P_1 D_{\lambda_i} P_1' = \frac{\lambda v}{k} P_1 P_1'.$$

By virtue of equations (2) and (4) we have

$$P_1 P_1' = I - \frac{1}{v} J$$

where J is a matrix of dimensions $v \times v$ in which every element is unity. Hence

$$C = \frac{\lambda v}{k}\left[I - \frac{1}{v} J\right].$$

Thus $\lambda_{i\alpha} = \lambda$ for all $i \neq \alpha$ hence, the result.

Acknowledgements: I would like to thank Dr. E. J. Williams and Dr. R. C. Bose for their suggestions and criticisms.

REFERENCES

[1] R. C. Bose, "Advanced Analysis of Variance and Covariance," Unpublished Notes.
[2] Sylvain Ehrenfeld, "On the Efficiency of Experimental Designs" *Ann. Math. Stat.*, Vol. 26 (1955), pp. 247–255.

[3] S. N. Roy, "A Report on some Aspects of Multivariate Analysis," North Carolina Institute of Statistics Mimeograph Series No. 121.
[4] H. W. Turnbull and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices*, Blackie and Son, Limited, London, 1952.
[5] K. D. Tocher, "The Design and Analysis of Block Experiments," JRSS Series B, Vol. 14 (1952), pp. 45-84.

A CHARACTERIZATION OF THE NORMAL DISTRIBUTION[1]

By J. N. K. Rao

Forest Research Institute, Dehra Dun, India

1. Introduction. Using characteristic functions Lukacs [3] has shown that a necessary and sufficient condition for the independence of the sample mean and variance is that the parent population be normal. Geisser [2] has derived a similar theorem concerning the sample mean and the first order mean square successive difference. In section 2 of this note a general theorem of which Lukacs' and Geisser's results are particular cases has been proved.

Lukacs [3] has extended his theorem to the multivariate case, namely, that a necessary and sufficient condition that the sample mean vector is distributed independently of the variance-covariance matrix is that the parent population be multivariate normal. In section 3, the general theorem of section 2 is extended to the multivariate population of which Lukacs' theorem for the multivariate population is a particular case. To prove the necessity of this theorem, we extend, to the multivariate case, Daly's [1] result that if $f(x)$ is the normal density, then the sample mean and $g(x_1 \cdots x_n)$ are independently distributed where $g(x_1 \cdots x_n) = g(x_1 + a, \cdots, x_n + a)$.

2. Univariate case. Let $x_1, \cdots, x_n$ be independent and identically distributed with density function $f(x)$ and mean μ and variance σ^2.

Let,

$$\bar{x} = n^{-1} \sum_{j=1}^{n} x_j \cdots \tag{2.1}$$

and

$$\delta^2 = \left(\sum_{t=1}^{m} \sum_{j=1}^{n} l_{tj}^2 \right)^{-1} \sum_{t=1}^{m} (l_{t1} x_1 + \cdots + l_{tn} x_n)^2, \qquad m \geqq 1 \tag{2.2}$$

where

$$\sum_{j=1}^{n} l_{tj} = 0 \quad \text{for} \quad t = 1, \cdots, m.$$

The following theorem is proved.

Received April 8, 1957; revised January 17, 1958.
[1] Supported by a Senior Research Training Scholarship from the Government of India.

THEOREM 1. *A necessary and sufficient condition that $f(x)$ be the normal density is that $\bar{x}$ and δ^2 are independent.*

PROOF. Following Lukacs [3] we derive the sufficiency. Now,

$$E(\delta^2) = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} l_{ij}^2\right)^{-1}\left\{\sum_i\sum_j l_{ij}^2 E(x_j^2) + \sum_{i=1}^{m}\sum_{j\neq j'}^{n} l_{ij}l_{ij'}E(x_j x_{j'})\right\}$$
$$= \sigma^2$$

The joint characteristic function of $\bar{x}$ and δ^2 is

$$\phi(t_1, t_2) = \int\int\cdots\int e^{it_1\bar{x}}e^{it_2\delta^2}f(x_1)\cdots f(x_n)\,dx_1\cdots dx_n .$$

Therefore

$$\frac{\partial}{\partial t_2}\phi(t_1, t_2)\,|_{t_2=0} = \phi_1(t_1)\,\frac{\partial}{\partial t_2}\phi_2(t_2)\,|_{t_2=0}, \tag{2.3}$$

where

$$\phi_1(t_1) = [\psi(t_1/n)]^n$$

and

$$\psi(t_1) = \int e^{it_1x}f(x)\,dx,$$

$$\begin{aligned}\frac{\partial}{\partial t_2}\phi(t_1, t_2)\,|_{t_2=0} &= i\,(\sum_i\sum_j l_{ij}^2)^{-1}\Big\{(\sum_i\sum_j l_{ij}^2)[\psi(t_1/n)]^{n-1}\int x^2e^{it_1x/n}f(x)\,dx \\ &\quad + 2\,(\sum_i\sum_{j\neq j'} l_{ij}l_{ij'})[\psi(t_1/n)]^{n-2}\Big[\int xe^{it_1x/n}f(x)\,dx\Big]^2\Big\} \\ &= i\Big\{[\psi(t_1/n)]^{n-1}\int x^2e^{it_1x/n}f(x)\,dx \\ &\quad - [\psi(t_1/n)]^{n-2}\Big[\int xe^{it_1x/n}f(x)\,dx\Big]^2\Big\},\end{aligned} \tag{2.4}$$

and

$$\frac{\partial}{\partial t_2}\phi_2(t_2)\,|_{t_2=0} = i\sigma^2.$$

Hence, Eq. (2.3) reduces to

$$-\psi(t)\,\frac{d^2\psi(t)}{dt^2} + \left[\frac{d\psi(t)}{dt}\right]^2 = \sigma^2[\psi(t)]^2, \tag{2.5}$$

the solution of which is the characteristic function of the normal distribution.

The necessary condition follows from Daly [1] who has proved that $\bar{x}$ and $g(x_1\cdots x_n)$ are independent in the normal case, if

$$g(x_1\cdots x_n) = g(x_1 + a, \cdots, x_n + a).$$

Since δ^2 is invariant under a translation, the theorem is proved.

In fact, the above result can easily be extended[2] to class of quadratic forms, namely those which are invaria and have non-zero expected values. For, Lukacs' metho when δ^2 is defined as follows:

$$(2.6) \qquad \delta^2 = \left(\sum_{t=1}^{m}\sum_{j=1}^{n} a_{tij}\right)^{-1}\left[\sum_{t=1}^{m}\sum_{i,j=1}^{n}\sum a_{tij}x_i x\right.$$

where $\sum_{j=1}^{n} a_{tij} = 0$ $(t = 1, \cdots, m,\ i = 1, \cdots, n)$, provi

$$\sum_{j=1}^{n} a_{tij} \neq 0 \ (t = 1, \cdots, m).$$

It will be noted that δ^2 defined in (2.2) above is a spe in (2.6) by putting $a_{tij} = l_{ti}l_{tj}$

Particular Cases.

(a) To obtain Lukacs' result, put

$$l_{tj} = 1 - \frac{1}{n} \quad \text{for} \quad t = j$$

$$= \frac{-1}{n} \quad \text{for} \quad t \neq j$$

$$\text{and } m = n.$$

(b) To get Geisser's result, put

$$l_{tj} = 1 \quad \text{when} \quad j = t + k$$

$$= -1 \quad \text{when} \quad j = t$$

$$= 0 \text{ for other values of } j$$

$$\text{and } m = n - k.$$

(c) An interesting extension of Geisser's result is: a nec condition for the independence of the sample mean and *an* successive difference is that the parent population be norm

The rth order mean square successive difference is given

$$\delta_r^2 = (n-r)^{-1}\left[\binom{r}{0}^2 + \cdots + \binom{r}{r-1}^2 + \binom{r}{r}^2\right]^{-1} \sum^{n-}$$

where

$$\Delta^r x_i = \binom{r}{0}x_{i+r} - \binom{r}{1}x_{i+r-1} + \cdots$$

To get the above result, put

$$l_{ij} = (-1)^{i+r-j}\binom{r}{i+r-j} \quad \text{when}$$

$$= 0 \quad \text{when} \quad 1 \leqq j \leqq i - 1$$

$$\text{and} \quad m = n - r.$$

---

[2] I am indebted to the referee for pointing this

3. Multivariate case. The same reasoning applies also to the multivariate case. Denote by $x_{\alpha i}$ ($\alpha = 1, \cdots, n$; $i = 1, \cdots, p$) the α observation on the ith variate, by $\bar{x}_i$, the sample mean of the ith variate,

(3.1) $$\delta_{ij} = \left[\left(\sum_{t=1}^{m} \sum_{\alpha=1}^{n} l_{t\alpha}^2\right)\right]^{-1} \sum_{t=1}^{m} \left\{\sum_{\alpha,\alpha'}^{n} l_{t\alpha} l_{t\alpha'} x_{\alpha i} x_{\alpha' j}\right\},$$

or more generally,

(3.2) $$\delta_{ij} = \left[\left(\sum_{t=1}^{m} \sum_{\alpha=1}^{n} a_{t\alpha\alpha}\right)\right]^{-1} \sum_{t=1}^{m} \left\{\sum_{\alpha,\alpha'}^{n} a_{t\alpha\alpha'} x_{\alpha i} x_{\alpha' j}\right\} \qquad (i, j = 1, \cdots, p),$$

where $\sum_{\alpha'=1}^{n} a_{t\alpha\alpha'} = 0$ ($t = 1, \cdots, m$; $\alpha = 1, \cdots, n$), provided

$$\sum_{\alpha=1}^{n} a_{t\alpha\alpha} \neq 0 \ (t = 1, \cdots, m).$$

Assuming that the distribution of $[\delta_{ij}]_{p\times p}$ is independent of the joint distribution of the p sample means $(\bar{x}_1, \cdots, \bar{x}_p)$ one obtains the equation,

(3.3) $$\frac{\psi_{ij}}{\psi} - \frac{\psi_i \psi_j}{\psi^2} = -\lambda_{ij},$$

where λ_{ij} is population covariance of the variates x_i and x_j,

$$\psi = \psi(t_1, \cdots, t_p) = \int\int \cdots \int e^{\iota(t_1 x_1 + \cdots + t_p x_p)} f(x_1 \cdots x_p)\, dx_1 \cdots dx_p.$$

$$\psi_i = \frac{\partial \psi}{\partial t_i}, \qquad \psi_{ij} = \frac{\partial^2 \psi}{\partial t_i \partial t_j}$$

If (3.3) is true for $i, j = 1, \cdots, p$, one has a set of partial differential equations which leads to the characteristic function to the multivariate normal distribution.

To prove the necessity, we give an extension of Daly's [1] lemma of which it is a particular case.

THEOREM 2. *Let* $g_l(x_{11}, \cdots, x_{n1}; \cdots; x_{1p}, \cdots, x_{np})$, $l = 1, \cdots, r$, *be functions of* $(x_{11}, \cdots, x_{n1}); \cdots, (x_{1p}, \cdots, x_{np})$ *and are such that*

$$g_l(x_{11} + a_1, \cdots, x_{n1} + a_1; \cdots; x_{1p} + a_p, \cdots, x_{np} + a_p) = g_l(x_{11}, \cdots, x_{n1}; \cdots; x_{1p}, \cdots, x_{np}).$$

The sample means $(\bar{x}_1, \cdots, \bar{x}_p)$ *are independently distributed of these* r *functions iff* $f(x_1 \cdots x_p)$ *has a* p*-variate normal distribution.*

PROOF. The joint characteristic function is

$$\phi(t_1, \cdots, t_p; \xi_1, \cdots, \xi_r) = \frac{1}{(2\pi)^{np/2} |\lambda_{ij}|^{n/2}} \int \cdots \int \exp\left\{\iota \sum_{\alpha=1}^{n} t_i x_{\alpha i}/n\right\} \exp\left\{\iota \sum_{l=1}^{r} \xi_l g_l\right\} \exp\left\{-\frac{1}{2} \sum_{\alpha=1}^{n} \sum_{i,j=1}^{p} \lambda^{ij} x_{\alpha i} x_{\alpha j}\right\} \times \prod_{\alpha=1}^{n} [dx_{\alpha 1} \cdots$$

where $(\iota)^2 = -1$.

Make the contragradient transformation

$$x_{\alpha i} = \sum_{j=1}^{p} c_{ij}\, y_{\alpha j}, \qquad t_i = \sum_{j=1}^{p} c_{ij}\, u_j \qquad i = 1, \cdots, p;\ \alpha = 1, \cdots, n.$$

Then,

$$\phi(t_1, \cdots, t_p; \xi_1, \cdots, \xi_r)$$
$$= \frac{1}{(2\pi)^{np/2} \mid \lambda_{ij} \mid^{n/2}} \int \cdots \int \exp\left\{\iota \sum_{\alpha=1}^{n} \sum_{i=1}^{p} u_i\, y_{\alpha i}/n\right\} \exp\left\{\iota \sum_{l=1}^{r} \xi_l\, g_l'\right\}$$
$$\exp\left\{-\frac{1}{2} \sum_{\alpha=1}^{n} \sum_{i=1}^{p} y_{\alpha i}^2/\rho_i\right\} \times \prod_{\alpha=1}^{n} [dy_{\alpha 1} \cdots dy_{\alpha p}],$$

where $\rho_1, \cdots, \rho_p$ are latent roots of the variance-covariance matrix and

$$g_l'(y_{11} + a_1, \cdots, y_{n1} + a_1; \cdots; y_{1p} + a_p, \cdots, y_{np} + a_p)$$
$$= g_l'(y_{11}, \cdots, y_{n1}; \cdots; y_{1p}, \cdots, y_{np}).$$

Put

$$\frac{y_{\alpha i}}{\sqrt{\rho_i}} - \frac{\iota u_i \sqrt{\rho_i}}{n} = Z_{\alpha i};$$

then

$$\phi(t_1, \cdots, t_p; \xi_1, \cdots, \xi_r)$$
$$= \frac{1}{(2\pi)^{np/2}} \int \cdots \int \exp\left\{-\frac{1}{2n} \sum_{i,j=1}^{p} \lambda_{ij}\, t_i\, t_j\right\} \exp\left\{\iota \sum_{l=1}^{r} \xi_l\, g_l''\right\}$$
$$\exp\left\{-\frac{1}{2} \sum_{\alpha=1}^{n} \sum_{i=1}^{p} Z_{\alpha i}^2\right\} \times \prod_{\alpha=1}^{n} [dZ_{\alpha 1} \cdots dZ_{\alpha p}],$$

where

$$g_l'' = g_l'(Z_{11}\sqrt{\rho_1}, \cdots, Z_{n1}\sqrt{\rho_1}; \cdots; Z_{1p}\sqrt{\rho_p}, \cdots, Z_{np}\sqrt{\rho_p})$$

and hence is a function of $(Z_{11}, \cdots, Z_{n1}); \cdots; (Z_{1p}, \cdots, Z_{np})$ only. Therefore,

$$\phi(t_1, \cdots, t_p; \xi_1, \cdots, \xi_r)$$
$$= \exp\left\{-\frac{1}{2n} \sum_{i,j=1}^{p} \lambda_{ij}\, t_i\, t_j\right\} \times (\text{a function of } \xi_1, \cdots, \xi_r \text{ only}).$$

Hence the theorem.

Particular case. The sample mean vector $(\bar{x}_1 \cdots \bar{x}_p)$ is independently distributed of products moments of any order if $f(x_1 \cdots x_p)$ has a p-variate normal density.

4. Acknowledgement. My thanks are due to Dr. K. R. Nair for his valuable suggestions and to the referee for his helpful comments.

REFERENCES

[1] J. F. DALY, "On the use of the sample range in [illegible]," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 71–74.
[2] S. GEISSER, "A note on the normal distribution," [illegible] pp. 858–859.
[3] E. LUKACS, "A characterization of the normal [illegible] Vol. 13 (1942), pp. 91–93.

A NOTE ON P.B.I.B. [illegible]

BY W. A. T[illegible]

University of [illegible]

Summary. The [illegible] P.B.I.B. [illegible] plete block design with m associative [illegible]

It is found that the C matrix of a P.B.I.B. [illegible] function of $m + 1$ [illegible] feels that this decomposition may be of [illegible] of P.B.I.B. designs.

1. The C matrix of a P.B.I.B. design. [illegible] of partially balanced designs, [illegible] for example, Bose and [illegible] [3].

The matrix

[illegible]

where

[illegible]

[illegible]

is of special [illegible]

In the case of a P.B.I.B. [illegible] form. We may write

(1.1) [illegible]

where B_s = [illegible] cording as the [illegible] $B_2, \cdots, B_m$ [illegible] (i, j)th [illegible] others. [illegible] of treatment [illegible] of treatment [illegible]

Received [illegible]

h associates, then this is the definition of p_{ss}^r. Note further that if $j = l$ $\sum_i [b_{ij}^{(s)}]^2 = n_s$. Thus

$$B_s B_s = [\sum_i b_{ij}^{(s)} b_{il}^{(s)}] = [\sum_i b_{ij}^{(s)} b_{il}^{(s)}]$$

$$= n_s I + \sum_i p_{ss}^i B_i .$$

rly

$$B_s B_t = [\sum_i b_{ij}^{(s)} b_{il}^{(t)}]$$

$$= \sum_i p_{st}^i B_i .$$

sider the equations

$$C = r(1 - 1/k)I - 1/k \sum_{i=1}^{m} \lambda_i B_i ,$$

$$CB_j = r(1 - 1/k)B_j - 1/k \sum_{i=1}^{m} \lambda_i B_i B_j , \qquad j = 1 \cdots m,$$

$$= r(1 - 1/k)B_j - 1/k \sum_{i \neq j} \lambda_i \left(\sum_{s=1}^{m} p_{ij}^s B_s \right)$$

$$- \lambda_j/k \left(n_j I + \sum_{s=1}^{m} p_{jj}^s B_s \right)$$

$$= - \frac{n_j \lambda_j}{k} I + \left[r(1 - 1/k) - 1/k \sum_i \lambda_i p_{ij}^j \right] B_j$$

$$- \sum_{s \neq j} 1/k \sum_i \lambda_i p_{ij}^s B_s \qquad j = 1 \cdots m .$$

rewrite these equations as

$$\begin{aligned} C &= d_{00} I + d_{01} B_1 + \cdots + d_{0m} B_m , \\ CB_1 &= d_{10} I + d_{11} B_1 + \cdots + d_{1m} B_m , \\ &\vdots \\ CB_m &= d_{m0} I + d_{m1} B_1 + \cdots + d_{mm} B_m , \end{aligned}$$

$$d_{00} = r(1 - 1/k),$$

$$d_{0i} = - \lambda_i/k, \qquad i = 1 \cdots m,$$

$$d_{j0} = - \frac{n_j \lambda_j}{k}, \qquad j = 1 \cdots m.$$

$$d_{jj} = r(1 - 1/k) - 1/k \sum_i \lambda_i p_{ij}^j , \qquad j = 1 \cdots m,$$

$$d_{js} = - 1/k \sum_i \lambda_i p_{ij}^s , \qquad s = 1 \cdots m; j \neq s.$$

itrary, and I is a $r \times r$ matrix, then by subtracting eI from C in (1.4)

we get the single matrix equation:

$$\begin{bmatrix} C - eI & & & 0 \\ & C - eI & & \\ & & \ddots & \\ 0 & & & C - eI \end{bmatrix} \begin{bmatrix} I \\ B_1 \\ \vdots \\ B_m \end{bmatrix}$$

$$= \begin{bmatrix} (d_{00} - e)I & d_{01} I & \cdots & d_{0m} I \\ d_{10} I & (d_{11} - e)I & \cdots & d_{1m} I \\ \cdots & \cdots & \cdots & \cdots \\ d_{m0} I & d_{m1} I & \cdots & (d_{mm} - e)I \end{bmatrix} \begin{bmatrix} I \\ B_1 \\ \vdots \\ B_m \end{bmatrix}.$$

Let D be the $(m + 1) \times (m + 1)$ square matrix:

$$D = \begin{bmatrix} d_{00} & d_{01} & \cdots & d_{0m} \\ d_{10} & d_{11} & \cdots & d_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ d_{m0} & d_{m1} & \cdots & d_{mm} \end{bmatrix}. \tag{1.7}$$

We could at this point use the B matrices to verify the following result:

THEOREM 1. *If e is a characteristic root of C then it is a characteristic root of D, and conversely if e is a characteristic root of D then it is a characteristic root of C.*

However, this theorem also follows from Lemma 3.1 of Connor and Clatworthy [3].

Using the matrices M and A of Lemma 3.1, with $z = kx - r(k - 1)$ we have

$$| M/k | = | xI - C |,$$

and

$$x| A/k | = | xI - D |.$$

This second relation follows by first adding all other rows of $| xI - D |$ to the first row and then subtracting the first column from all others. Theorem 1 then follows from Connor and Clatworthy's lemma.

2. The principal idempotent matrices of C. (If the reader is unfamiliar with the properties of principal idempotent matrices, then he may consult [4].) Let e be a characteristic root of C, and let $E(e)$ be the principal idempotent matrix of C corresponding to e. Theorem 1 then states that e is a root of D. B_0 will denote the identity matrix.

THEOREM 2. *$E(e) = \sum_{i=0}^{m} c_i B_i$, where $(c_0, c_1, \cdots, c_m)$ is a characteristic vector of D corresponding to e.*

PROOF. $E(e)$ must be a polynomial in C. Therefore, $E(e) = \sum_{i=0}^{m} c_i B_i$ according to (1.1), (1.2), and (1.3). At this point in the proof $c_0, c_1, \cdots, c_m$ are arbitrary constants. Now, $E(e)(C - eI) = 0$ since this is a property of principal idempotent matrices for C real and symmetric.

We rewrite this relation

$$(c_0 I, c_1 I, \cdots, c_m I) \begin{bmatrix} C - eI & & & 0 \\ & C - eI & & \\ & & \ddots & \\ 0 & & & C - eI \end{bmatrix} \begin{bmatrix} I \\ B_1 \\ \vdots \\ B_m \end{bmatrix} = 0. \tag{2.1}$$

Using 1.6 and the linear independence of the B's, 2.1 yields

$$(2.2)\quad (c_0 I, c_1 I, \cdots, c_m I)\begin{bmatrix} (d_{00}-e)I & d_{01} I & \cdots & d_{0m} I \\ d_{10} I & (d_{11}-e)I & & d_{1m} I \\ \vdots & \vdots & & \vdots \\ d_{m0} I & d_{m1} I & \cdots & (d_{mm}-e)I \end{bmatrix} = 0.$$

Therefore

$$(2.3)\qquad (c_0, c_1, \cdots, c_m)(D - eI) = 0.$$

If C has m^* distinct non-zero characteristic roots, $e_1, e_2, \cdots, e_{m^*}$, then we may write

$$C = e_1 E(e_1) + e_2 E(e_2) + \cdots + e_{m^*} E(e_{m^*}).$$

Now using Theorem 2 we have

THEOREM 3. *The C matrix of* a P.B.I.B. (m) *may be expressed as a linear function of the $m + 1$ commutative and linearly independent matrices* $B_0, B_1, \cdots, B_m$.

REFERENCES

[1] R. C. BOSE, *Least Square Aspects of Analysis of Variance*, Institute of Statistics, University of North Carolina.
[2] R. C. BOSE AND T. SHIMAMOTO, "Classification and Analysis of Partially Balanced Incomplete Block Designs with Two Associate Classes," *J. Amer. Stat. Assn.*, Vol. 47, No. 258.
[3] W. S. CONNOR AND W. H. CLATWORTHY, "Some Theorems for Partially Balanced Designs," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 100–113.
[4] S. PERLIS, *Theory of Matrices*, Addison-Wesley, Mass., 1952, Theorem 9-5, p. 175.

ON A FACTORIZATION THEOREM IN THE THEORY OF ANALYTIC CHARACTERISTIC FUNCTIONS[1]

Dedicated to Paul Lévy on the occasion of his seventieth birthday

BY R. G. LAHA

The Catholic University of America

1. Introduction. Let $F(x)$ be a distribution function, that is, a non-decreasing right-continuous function such that $F(-\infty) = 0$ and $F(+\infty) = 1$. The characteristic function

$$(1.1)\qquad \phi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)$$

of the distribution function $F(x)$ is defined for all real t. A characteristic function is said to be an *analytic characteristic function* if it coincides with a regular analytic function $\phi(z)$ in some neighborhood of the origin in the complex z-plane.

Received November 22, 1957; revised March 3, 1958.
[1] This work was supported by the National Science Foundation through grant NSF-G-4220.

Then it follows from a theorem due to Boas [1] that the analytic characteristic function $\phi(z)$ is also regular in a horizontal strip $-\alpha < Im\, z < +\beta$ of the complex z-plane containing the real axis. It is also well known that the analyticity of the characteristic function $\phi(z)$ in the horizontal strip $| Im\, z | \leqq R (R > 0)$ is equivalent to the condition that (a) the corresponding distribution function $F(x)$ has moments μ_k of all orders k and further (b) $\lim \sup_{k\to\infty}[\mu_k/k!]^{1/k}$ is finite and equal to $1/R$. In other words, the analytic characteristic function $\phi(z)$ has the power series expansion

$$(1.2) \qquad \phi(z) = \sum_{k=0}^{\infty} \frac{i^k \mu_k}{k!} z^k$$

about the origin $z = 0$ in the circle $| z | \leqq R$ (z complex) where $R > 0$ is the radius of convergence of the series. The characteristic function $\phi(z)$ is said to be an *entire characteristic function* if its strip of regularity comprises the whole complex z-plane. A summary of most of the important properties of analytic characteristic functions is given in a recent paper by Lukacs [6].

In the present paper we shall discuss some results concerning the decomposition properties of analytic characteristic functions. In this direction a very interesting theorem has been recently obtained by Linnik [5], [7] which may be considered as an analytical extension of Cramér's theorem [2] on the normal law. The theorem is as follows:

THEOREM OF LINNIK. *Let $\phi_1(t), \phi_2(t), \cdots, \phi_n(t)$ denote the characteristic functions of some non-degenerate distributions and let $\alpha_1, \alpha_2, \cdots, \alpha_n$ be some positive numbers. Let the functions $\phi_j(t)$ satisfy the equation*

$$(1.3) \qquad \prod_{j=1}^{n} \{\phi_j(t)\}^{\alpha_j} = \exp \{i\mu t - \tfrac{1}{2}\sigma^2 t^2\}$$

for all real t in a certain neighborhood $| t | < \delta (\delta > 0)$ of the origin, where $\sigma^2 > 0$ and μ are real numbers. Then each factor $\phi_j(t)$ is the characteristic function of a normal distribution.

In the following section we shall deal with some related factorization theorems (Theorems 2.1 and 2.2) for analytic characteristic functions. These theorems may be considered as generalizations of the theorem of Linnik stated above.

2. The Theorems. We now consider the following theorems:

THEOREM 2.1. *Let $\phi_1(t), \phi_2(t), \cdots, \phi_n(t)$ denote the characteristic functions of some non-degenerate distributions. Let further $\phi(z)$ denote an analytic characteristic function and $\alpha_1, \alpha_2, \cdots, \alpha_n$ be some positive numbers. Let the functions $\phi_j(t)$ satisfy the equation*

$$(2.1) \qquad \prod_{j=1}^{n} \{\phi_j(t)\}^{\alpha_j} = \phi(t)$$

for all real t in a certain neighborhood $|t| < \delta (\delta > 0)$ of the origin. Then each of the factors $\phi_j(z)$ is also an analytic characteristic function which is regular at least in the strip of regularity of $\phi(z)$.

This theorem has already been obtained by Dugué and stated without proof

and without any zeros. Hence applying Theorems 2.1 and 2.2, it follows at once that each of the factors $\phi_j(z)$ is also an entire characteristic function of order not exceeding two and without any zeros in the complex plane. Then the proof follows at once, using the factorization theorem of Hadamard to each of the factors $\phi_j(z)$.

In conclusion the author wishes to express his thanks to Professor Eugene Lukacs for calling his attention to the paper by Dugué [3].

REFERENCES

[1] R. P. Boas, "Sur les séries et intégrales de Fourier à coefficients positifs," *C.R. Acad. Sci. Paris*, Vol. 228 (1949), pp. 1837–1838.

[2] H. Cramér, "Über eine Eigenschaft der normalen Verteilungsfunktion," *Math. Zeit.*, Vol. 41 (1936), pp. 405–414.

[3] D. Dugué, "Résultats sur les fonctions absolument monotones et applications à l'arithmétique de fonctions de type positif," *C.R. Acad. Sci. Paris*, Vol. 244 (1957), pp. 715–717.

[4] R. G. Laha, "On a characterization of the normal distribution from properties of suitable linear statistics," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 126–139.

[5] Yu. V. Linnik, "A problem on characteristic functions of probability distributions (Russian)," *Uspehi Matem. Nauk.*, Vol. 10 (1955), pp. 137–138.

[6] E. Lukacs "Les fonctions caractéristiques analytiques," *Ann. Inst. Hen. Poincaré* Vol. 15 (1957), pp. 217–251.

[7] A. A. Zinger and Yu. V. Linnik, "On an analytical extension of a theorem of Cramér and its application," (Russian), *Vestnik Leningrad Univ.*, Vol. 10 (1955), pp. 51–56.

BOUNDS FOR MILLS' RATIO FOR THE TYPE III POPULATION

By A. V. Boyd

University of the Witwatersrand

1. Introduction and summary. Cohen [1] and Des Raj [2] have shown that in estimating the parameters of truncated type III populations, it is necessary to calculate for several values of x the Mills ratio of the ordinate of the standardized type III curve at x to the area under the curve from x to ∞. Des Raj [3] has also noted that for large values of x the existing tables of Salvosa [4] are inadequate for this purpose and he has found lower and upper bounds for the ratio. The object of this note is to improve these bounds, by obtaining monotonic sequences of lower and upper bounds through the use of continued fractions.

2. Approximations to the ratio. Taking the type III population in the standardized form

$$C\,f(x)\,dx, \qquad -2/\alpha \leqq x \leqq \infty, \qquad 0 \leqq \alpha \leqq 2,$$

Received August 5, 1957.

where

$$f(x) = \left(1 + \frac{\alpha x}{2}\right)^{(4/\alpha^2)-1} e^{-2x/\alpha}$$

and

$$C = (4/\alpha^2)^{(4/\alpha^2)-1/2} e^{-4/\alpha^2} [\Gamma(4/\alpha^2)]^{-1},$$

Des Raj [3] puts

$$G(x) = \int_x^\infty f(t)\,dt \qquad \text{and} \qquad \mu(x) = f(x)/G(x)$$

and obtains

$$\frac{2x + \alpha}{\alpha x + 2} \leqq \mu(x) \leqq \frac{2}{(x^2 + 2\alpha x + 4)^{1/2} - x}.$$

However, by making the substitution $\alpha^2 v = 2(\alpha t + 2)$ in the integral for $G(x)$, we find

$$G(x) = e^a a^{1/2-a} \int_X^\infty e^{-v} v^{a-1}\,dv,$$

where

$$a = 4/\alpha^2 \qquad \text{and} \qquad X = a + a^{1/2} x.$$

Now, by Wall [5] equation (92.9),

$$\int_X^\infty e^{-v} v^{a-1}\,dv = e^{-X} X^a \left\{ \frac{1}{X+} \, \frac{1-a}{1+} \, \frac{1}{X+} \, \frac{2-a}{1+} \, \frac{2}{X+} \, \frac{3-a}{1+} \, \frac{3}{X+} \cdots \right\}$$

for all a if $X > 0$. On substituting and simplifying it is then found that for $x > -2/\alpha$,

$$1/\mu(x) = a^{-1/2} X \left\{ \frac{1}{X+} \, \frac{1-a}{1+} \, \frac{1}{X+} \, \frac{2-a}{1+} \, \frac{2}{X+} \, \frac{3-a}{1+} \, \frac{3}{X+} \cdots \right\}$$

The approximants to the continued fraction on the righthand side lead to approximations to $\mu(x)$. The first seven of these are

$$\mu_1(x) = 2/a, \qquad \mu_2(x) = \frac{2x + \alpha}{\alpha x + 2}, \qquad \mu_3(x) = \frac{4(x + \alpha)}{2\alpha x + \alpha^2 + 2},$$

$$\mu_4(x) = \frac{2(2x^2 + 4\alpha x + \alpha^2 + 2)}{(\alpha x + 2)(2x + 3\alpha)},$$

$$\mu_5(x) = \frac{2(2x^2 + 6\alpha x + 2 + 3\alpha^2)}{2\alpha x^2 + (5\alpha^2 + 4)x + 10\alpha + \alpha^3},$$

$$\mu_6(x) = \frac{2(4x^3 + 18\alpha x^2 + 6(2 + 3\alpha^2)x + 3\alpha^3 + 14\alpha)}{(\alpha x + 2)(4x^2 + 16\alpha x + 11\alpha^2 + 8)},$$

$$\mu_7(x) = \frac{8(x^3 + 6\alpha x^2 + 3(3\alpha^2 + 1)x + 3\alpha^3 + 5\alpha)}{4\alpha x^3 + 2(11\alpha^2 + 4)x^2 + 26\alpha(\alpha + 2)x + 3\alpha^4 + 52\alpha^2 + 16}$$

It should be noted that $\mu_2(x)$ is Des Raj's lower bound for $\mu(x)$. By elementary algebra it can be shown that $\mu_3(x)$ exceeds $\mu_2(x)$ for all relevant α and x, an

TABLE I

Values of $\mu_r(x)$ when $a = 4$, $\alpha = 1$

x	$\mu_2(x)$	$\mu_3(x)$	$\mu_6(x)$	$\mu_7(x) = \mu(x)$	$\mu_5(x)$	$\mu_4(x)$	$\frac{2}{(x^2+2\alpha x+4)^{\frac{1}{2}} - x}$
−.50	0.000	0.500	0.667	0.692	0.714	1.000	0.869
.00	0.500	0.800	0.894	0.901	0.909	1.000	1.000
.50	0.800	1.000	1.057	1.059	1.062	1.100	1.117
1.00	1.000	1.143	1.180	1.180	1.182	1.200	1.215
1.50	1.143	1.250	1.275	1.275	1.276	1.286	1.298
2.00	1.250	1.330	1.351	1.351	1.351	1.357	1.366
2.50	1.330	1.400	1.413	1.413	1.413	1.417	1.423
3.00	1.400	1.455	1.464	1.464	1.464	1.467	1.472
3.50	1.455	1.500	1.507	1.507	1.508	1.509	1.513
4.00	1.500	1.538	1.544	1.544	1.545	1.545	1.549

TABLE II

Values of $\mu_r(x)$ when $a = 16/9$, $\alpha = 1.5$

x	$\mu_2(x)$	$\mu_4(x)$	$\mu_3(x)$	$\mu_1(x)$
−.50	0.400	0.800	0.842	1.333
.00	0.750	0.944	0.960	1.333
.50	0.909	1.024	1.032	1.333
1.00	1.000	1.076	1.081	1.333
1.50	1.059	1.114	1.116	1.333
2.00	1.100	1.141	1.143	1.333
2.50	1.131	1.162	1.163	1.333
3.00	1.154	1.180	1.180	1.333
3.50	1.172	1.193	1.194	1.333
4.00	1.188	1.205	1.206	1.333

Further, for $x = 0$, $\mu_5 = 0.9523$, $\mu_6 = 0.9504$, and $\mu_7 = 0.9515$.

hat, for all relevant α and for $x > \max(0, 2/\alpha - 2\alpha)$, $\mu_4(x)$ is less than Des Raj's upper bound.

3. Convergence of the approximants for integral a. We suppose henceforth hat $x > -\alpha/2$. (All the inequalities to be derived appear to hold over at least part of the range $-2/\alpha \leqq x \leqq -\alpha/2$, but as we are interested only in large positive x we shall not worry to extend their range of validity.) If $a = n$ then $X + i - a = (2x + i\alpha)/\alpha > 0$ for $i = 1, 2, 3, \cdots$ and

$$i - a\begin{cases} < 0 & \text{for } i = 1 \text{ to } n - 1, \\ = 0 & \text{for } i = n. \end{cases}$$

Hence, by considering the approximants to

$$\frac{1}{X+}\,\frac{1-a}{1+}\,\frac{1}{X+}\,\frac{2-a}{1+}\,\frac{2}{X+}\cdots$$

it is easily verified that $\mu_1(x), \mu_2(x), \cdots, \mu_{2n-1}(x)$ satisfy the inequalities

$$\mu_2 < \mu_3 < \mu_6 < \mu_7 < \cdots < \mu < \cdots < \mu_9 < \mu_8 < \mu_5 < \mu_4 < \mu_1 .$$

$\mu_{2n-1}(x)$ is of course equal to $\mu(x)$ since the $(2n)$th partial numerator of the continued fraction vanishes. The rapidity of the convergence of the sequence $\mu_r(x)$ in the case $a = 4$ is indicated by Table I, where Des Raj's numerical bounds [3] are included for comparison.

4. Convergence of the approximants for non-integral a. If $n < a < n + 1$ then $X + i - a > 0$ for $i = 1, 2, \cdots$ and

$$i - a \begin{cases} <0 & \text{for } i = 1 \text{ to } n, \\ >0 & \text{for } i = n + 1, n + 2, \cdots, \end{cases}$$

so that $\mu_1(x), \cdots, \mu_{2n}(x)$ satisfy the same inequalities as in the case of integral a, while $\mu_{2n-1}(x), \mu_{2n+1}(x), \mu_{2n+3}(x), \cdots$ and $\mu_{2n}(x), \mu_{2n+2}(x), \mu_{2n+4}(x), \cdots$ form monotonic sequences approaching $\mu(x)$, one from above and the other from below. Thus if $2r - 1 < a < 2r$ then we have

$$\begin{aligned} \mu_2 < \mu_3 < \mu_6 < \mu_7 < \cdots < \mu_{4r-10} < \mu_{4r-9} < \mu_{4r-6} < \mu_{4r-5} \\ < \mu_{4r-2} < \mu_{4r} < \mu_{4r+2} < \cdots < \mu < \cdots < \mu_{4r+1} < \mu_{4r-1} < \mu_{4r-3} \\ < \mu_{4r-4} < \mu_{4r-7} < \mu_{4r-8} < \cdots < \mu_9 < \mu_8 < \mu_5 < \mu_4 < \mu_1 \end{aligned}$$

and if $2r < a < 2r + 1$ then

$$\begin{aligned} \mu_2 < \mu_3 < \mu_6 < \mu_7 < \cdots < \mu_{4r-6} < \mu_{4r-5} < \mu_{4r-2} < \mu_{4r-1} \\ < \mu_{4r+1} < \mu_{4r+3} < \cdots < \mu < \cdots < \mu_{4r+4} < \mu_{4r+2} < \mu_{4r} < \mu_{4r-3} \\ < \mu_{4r-4} < \mu_{4r-7} < \mu_{4r-8} < \cdots < \mu_9 < \mu_8 < \mu_5 < \mu_4 < \mu_1 . \end{aligned}$$

Table II indicates the rapidity of the convergence of μ_r in the case $a = 16/9$

REFERENCES

[1] A. C. Cohen, "Estimating parameters of Pearson type III populations from truncated samples," *J. Amer. Stat. Assn.*, Vol. 45 (1950), pp 411–423.
[2] Des Raj, "On estimating the parameters of type III populations from truncated samples," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp 336–349.
[3] Des Raj, "On Mills' ratio for the type III population," *Ann. Math Stat.*, Vol. 24 (1953) pp. 309–312.
[4] L. R Salvosa, "Tables of Pearson's Type III function," *Ann. Math Stat*, Vol. 1 (1930), appended.
[5] H S Wall, *Continued Fractions*, Van Nostrand, 1948

ON THE COMMUTATIVITY OF OPERATORS IN STOCHASTIC MODELS FOR LEARNING[1]

By Manfred Kochen

Harvard University[2]

Introduction. Bush and Mosteller[3] have shown that a very fruitful model for the analysis of certain experiments on Learning in animals can be developed in terms of linear operators, Q, which are defined as follows:

$$Qp = \alpha p + (1 - \alpha)\lambda \qquad 0 \leqq p \leqq 1, \quad 0 \leqq Qp \leqq 1.$$

The probability (measured as the relative frequency over a number of supposedly identical animals) that an animal makes a certain one of two possible responses on the kth trial is denoted by p_k, to be substituted for p in the above equation. The two alternatives might be going to the right and to the left in a T-maze, and p_k might be the probability of going to the right. The variable $Q_i p_k$ represents the probability that the animal makes the proper response (e.g. going to the right) on the $k + 1$st trial after the occurrence of the ith of several possible events. It is often sufficient to consider only two events, E_1 and E_2 (e.g. reward and punishment) and their associated operators Q_1 and Q_2. The learning process is assumed to be described by the following recursive (Markov-type) relation:

$$p_{k+1} = Q_i p_k \equiv \alpha_i p_k + (1 - \alpha_i)\lambda_i \quad 0 \leqq p_k \leqq 1, \quad k = 0, 1, 2, \cdots$$

$$0 \leqq Q_i p_k \leqq 1 \qquad i = 1, 2 \qquad\qquad k = 0, 1, 2, \cdots$$

after event E_i has occurred. The parameters α_i, λ_i $i = 1, 2$ are to be statistically estimated in order to obtain a good fit between computed and observed data. If, for instance, the sequence of events $E_1E_2E_1E_2$ were to occur, then $p_4 = Q_2Q_1Q_2Q_1p_0$. The estimation of α_1, α_2, λ_1, λ_2, from even this 4-trial experiment presents considerable technical difficulties. If it were known, however, that the two operators commute, then $p_4 = Q_1^2Q_2^2p_0$, which simplifies the estimation problem considerably. If the operators do not commute, and nothing appears to indicate that they do in general, it might be inquired if there is not some function of p_k into $f(p_k)$ such that the induced operators on $f(p_k)$ will commute.

Results. Consider the closed unit interval [0, 1], and let p be any point in it.

Received March 12, 1957.

[1] This problem was suggested by Professor F. Mosteller.

[2] This work was done while the author was at Harvard University under a research grant from the Ford Foundation in the spring of 1956. He is now at the IBM Research Center.

[3] R. R. Bush and F. Mosteller, Stochastic Models for Learning, John Wiley and Sons, N. Y., 1955.

From the restriction that $0 \leqq Q_i p \leqq 1$, it is easily deduced[4] that $0 \leqq \lambda_i \leqq 1$, and

$$\operatorname*{Max}_k \frac{\lambda_k}{\lambda_k - 1} \leqq \alpha_i \leqq 1, \qquad i = 1, 2.$$

Let f be a continuous function on [0, 1]. Suppose that the operator Q_i on p induces a transformation T_i on $f(p)$ such that

$$f(Q_i p) = T_i f(p) \text{ for every } p \varepsilon [0, 1].$$

The question arises whether there exists an f with the above properties and such that

$$T_1 T_2 f(p) = T_2 T_1 f(p) \text{ for all } p \varepsilon [0, 1]$$

regardless of whether $Q_1 Q_2 p = Q_2 Q_1 p$. The following result answers this question:

THEOREM. $T_1 T_2 f(p) = T_2 T_1 f(p)$ *if and only if* f *is a periodic function with period* $(1 - \alpha_1)(1 - \alpha_2)(\lambda_1 - \lambda_2)$.

PROOF.

(a) Suppose that

$$T_1 T_2 f(p) = T_2 T_1 f(p).$$

Then

$$(T_1 T_2 - T_2 T_1) f(p) = 0.$$

Observe that

$$T_1 T_2 f(p) = T_1 f(Q_2 p) = f(Q_1 Q_2 p),$$

so that

$$(T_1 T_2 - T_2 T_1) f(p) = f(Q_1 Q_2 p) - f(Q_2 Q_1 p) = 0.$$

But

$$Q_1 Q_2 p = \alpha_1[\alpha_2 p + (1 - \alpha_2)\lambda_2] + (1 - \alpha_1)\lambda_1 = ap + b,$$

where

$$a = \alpha_1 \alpha_2 \text{ and } b = \alpha_1(1 - \alpha_2)\lambda_2 + (1 - \alpha_1)\lambda_1$$

and

$$Q_2 Q_1 p = \alpha_2[\alpha_1 p + (1 - \alpha_1)\lambda_1] + (1 - \alpha_2)\lambda_2 = ap + b'$$

where

$$b' = \alpha_2(1 - \alpha_1)\lambda_1 + (1 - \alpha_2)\lambda_2 .$$

---

[4] R. R. Bush, F. Mosteller and G. L. Thompson, "A Formal Structure for Multiple-Choice Situations", Decision Processes, Eds. Thrall, Coombs and Davis, J. Wiley and Sons, N. Y., 1954, Ch. VIII.

Hence

$$f(ap + b) - f(ap + b') = 0 \text{ for all } p \in [0, 1].$$

Let

$$q = ap + b \text{ so that } f(q) = f(q + (b - b')).$$

This defines a periodic function with period

$$\begin{aligned}\mu = b - b' &= \alpha_1(1 - \alpha_2)\lambda_2 + (1 - \alpha_1)\lambda_1 - \alpha_2(1 - \alpha_1)\lambda_1 - (1 - \alpha_2)\lambda_2 \\ &= (1 - \alpha_1)(\lambda_1 - \alpha_2\lambda_1) + (1 - \alpha_2)(\alpha_1\lambda_2 - \lambda_2) \\ &= (1 - \alpha_1)\lambda_1(1 - \alpha_2) + (1 - \alpha_2)\lambda_2(\alpha_1 - 1) \\ &= (1 - \alpha_1)(1 - \alpha_2)(\lambda_1 - \lambda_2).\end{aligned}$$

(b) Now suppose that $f(p) = f(p + \mu)$ for all $p \in [0, 1]$ and some μ. Then $f(Q_1Q_2p) - f(Q_2Q_1p) = 0$ only if $(Q_1Q_2 - Q_2Q_1)p = k\mu$, $k = 0, 1, 2 \cdots$. But

$$(Q_1Q_2 - Q_2Q_1)p = (1 - \alpha_1)(1 - \alpha_2)(\lambda_1 - \lambda_2) = k\mu.$$

Letting $k = 1$, μ has the same value as above, and $(T_1T_2 - T_2T_1)f(p) = 0$. QED. All the equal signs should be understood as identities.

COROLLARY 1. *If Q_1 and Q_2 commute, then $\mu = 0$. This clearly occurs if and only if: $\alpha_1 = 1$ or $\alpha_2 = 1$ or $\lambda_1 = \lambda_2$.*

COROLLARY 2. *If $0 \leqq \alpha_i \leqq 1$ and $0 \leqq \lambda_i \leqq 1$ then $|\mu| \leqq 1$ with $\mu = 1$ if $\alpha_1 = \alpha_2 = 0$ or $\lambda_1 = 0, \lambda_2 = 1$ or $\lambda_1 = 1, \lambda_2 = 0$.*

Suppose that Q_1 and Q_2 do not commute. It is then desirable that f can transform p_0 such that

$$Q_1Q_2p_0 = f^{-1}T_1T_2f(p_0) = f^{-1}T_2T_1f(p_0).$$

Clearly, since f is periodic, it will not have a single-valued inverse. However, if bounds on $Q_1Q_2p_0$ are known, $A \leqq Q_1Q_2p_0 \leqq B$, such that $B - A \leqq \mu/2$, it may be possible to recover $p_2 = Q_1Q_2p_0$. For experiments in which the probability of one response becomes eventually very high and that of the other very low $|\lambda_1 - \lambda_2| \cong 1$. If, in addition, the experiment is such that the event E_1 has the same effect on one response as the event E_2 has on the other, α_1 may be taken equal to α_2. Call the common value α. Finally, if it can be estimated that α does not exceed some number C (e.g. 1/2) then $\mu/2 = (1 - C)^2/2$. This bound is largest when $C \sim 0$, and this implies that $\mu \sim 1$, by the above corollary. In this case f may have a single-valued inverse. In general, to have a single-valued inverse f ought to be monotonic inside $[A, B]$ provided that

$$A \leqq p_k \leqq B \qquad k = 0, 1, 2, \cdots$$

For instance, if $\mu = 1/2$ and $f(p) = \sin(2\pi/1/2)p$, and $7/8 \leqq p_k \leqq 1$, $k = 0, 1, 2, \cdots$ then $f(p_k)$ has a single-valued inverse, and the commutativity of T_1 and T_2 can be utilized.

General Remarks. Consider the case where there are r instead of 2 response classes. Then it is convenient to regard the r probabilities $p_1, \cdots, p_r$ as a normalized column vector, p. With t possible events, there are t corresponding linear operators, which can be represented by t $r \times r$ stochastic matrices, $M_1, \cdots, M_i, \cdots, M_t$. Then, the value of the vector p at the $k + 1$st trial, after the occurrence of event E_i, is given by $M_i p_k$ where p_k is the value of the vector at the kth trial. Under the assumption of combining classes, T_i may be written as $M_i = \alpha_i I + (1 - \alpha_i)\Lambda_i$ where I is the $r \times r$ identity matrix, and Λ_i is an $r \times r$ matrix in which all columns are identical, and the r entries are denoted by $\lambda_1^{(i)}, \cdots, \lambda_r^{(i)}$. It is then readily shown that the commutator of M_i and M_j is the vector: $\mu = (1 - \alpha_i)(1 - \alpha_j)(\Lambda_i - \Lambda_j)^*$. The last term $(\Lambda_i - \Lambda_j)^*$ is any of the r identical column vectors of the matrix $(\Lambda_i - \Lambda_j)$. It is now necessary to find f such that $f(M_i p) = T_i f(p)$ and such that $T_i T_j f(p) = T_j T_i f(p)$, where $f(p)$ denotes the column vector with elements $f(p_1), \cdots, f(p_r)$. The theorem goes through as before, these conditions being satisfied if and only if f is periodic with $f(p) = f(p + \mu)$, where μ is the commutator vector defined above. The determination of conditions under which f has an inverse is a somewhat deeper question. For the present, it is sufficient to remark that if the qth component of p_k is bounded by A_q and B_q for some $q \leqq r$ and f is monotone in $[A_q, B_q]$, then f has an inverse in that region, and the values of this qth component on successive trials can be used to estimate the parameters.

Returning to the case of $r = 2$ and $t = 2$, it appears that for a given Q_1 and Q_2 half the commutator $\mu/2$, gives a measure of the largest set of values of p on which it is possible to find a 1–1 mapping f such that the induced transformations T_1 and T_2 commute. At the same time, μ also gives a measure of the fraction of the interval [0, 1] on which the commutativity of Q_1 and Q_2 fails to hold.

REFERENCE

[1] *Stochastic Models for Learning*, John Wiley and Sons, New York, 1955.

ADDENDA TO "INTRA BLOCK ANALYSIS FOR FACTORIALS IN TWO-ASSOCIATE CLASS GROUP DIVISIBLE DESIGNS"[1]

By Ralph Allan Bradley and Clyde Young Kramer

Virginia Polytechnic Institute

1. Nair and Rao [1] in a very fundamental paper discussed confounding in asymmetrical (asymmetrical in the factor levels) factorial experiments. They gave a general formulation of the combinatorial set-up for balanced confounded designs, assuming their existence, of asymmetrical factorial experiments and

Received April 7, 1958.

[1] Research sponsored by the Statistics Branch, Office of Naval Research.

showed how to construct some optimum designs for two-factor experiments with some extensions to three and four factors.

Requirements for balanced confounded designs of asymmetrical factorials were set forth. Using their notation, we let $(i_1, \cdots, i_m)$ be the treatment combination with the i_tth level of factor F_t, $t = 1, \cdots, m$, F_t having s_t levels. There are $v = \prod_t s_t$ treatment combinations to be arranged in b blocks of k experimental units with no treatment combination on two units of the same block. Requirements for balanced confounding were:

(i) Every treatment combination is replicated r times.

(ii) The treatments $(i_1, \cdots, i_m)$ and $(j_1, \cdots, j_m)$ occur together in $\lambda_{k_1, \cdots, k_m}$ blocks where $k_t = 0$ or 1 as $i_t = j_t$ or $i_t \neq j_t$.

Nair and Rao discussed two-factor experiments in detail showing the estimation of treatment differences, efficiency and amount of information, and tests of significance.

2. Nair [2] in a short paper in 1953 showed that the earlier work of Bose and Connor [3] on group divisible, partially balanced, incomplete block designs with two associate classes could be regarded as a special case of the analysis for confounded asymmetrical factorial experiments with two factors. Also, he showed that designs constructed by Nair and Rao correspond to designs of the semi-regular class of group divisible designs typed by Bose and Shimamoto [4].

3. Kramer and Bradley [5], using group divisible designs catalogued by Bose, Clatworthy, and Shrikhande [6], showed how factorial treatment combinations may be used in these designs and presented the straight-forward least squares derivation of the intra-block analysis for such experiments. This essentially completes the cycle. The discussion of confounding in asymmetrical factorials is the most general of the papers; the factors could be regarded as pseudo-factors to derive the analysis for non-factorial treatments in the two-associate class group divisible designs. Finally, the treatments in the group divisible designs were replaced by factorial treatment combinations to produce confounded asymmetrical factorials.

4. Analyses for the basic two-factor factorial in [5] could have been based on the work of Nair and Rao [1] and Nair [2]. The association of notation (the Bradley-Kramer notation followed by that of Nair and Rao), where notations differed, is as follows:

$$m, s_2 ; n, s_1 ; \lambda_1, \lambda_{10} ; \lambda_2, \lambda_{01} = \lambda_{11} ; \qquad (\lambda_1 + rk - r)/k, p_{11} = p_{1\cdot} ;$$

$$mn\lambda_2/k, p_{\cdot 1} ; \qquad Q_{ij}, Q(i, j);$$

$$t_{ij}, t(i, j); \qquad A\text{-factor}, F_2\text{-factor}; \qquad \text{and } C\text{-factor}, F_1\text{-factor}.$$

The association of notations leads to equivalences of results. In the order as before, Table 1 corresponds to Table 2, variances of effects in (5.22) and (5.23)

with (3.23) and (3.22), and efficiencies (5.27), (5.28), and (5.29) with those indicated on the bottom of page 113 of [1].

5. We are indebted to K. R. Nair for drawing these matters to our attention.

REFERENCES

[1] K R. Nair and C. R. Rao, "Confounding in asymmetrical factorial experiments," *J. Roy Stat Soc. B*, Vol. 10 (1948), pp 109–131.
[2] K R. Nair, "A note on group divisible incomplete block designs," *Calcutta Stat. Assoc. Bull*, Vol 5 (1953), pp. 30–35.
[3] R C Bose and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann Math Stat*, Vol. 23 (1952), pp 367–383
[4] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced designs with two associate classes," *J. Amer Stat Assn*, Vol. 47 (1952), pp. 151–190
[5] C. Y. Kramer and R. A. Bradley, "Intra-block analysis for factorials in two-associate class group divisible designs," *Ann Math. Stat*, Vol 28 (1957), pp 349–361.
[6] R. C Bose, W. H. Clatworthy, and S S. Shrikhande, "Tables of partially balanced designs with two associate classes," *Tech. Bull. No 107 (1954), North Carolina Agricultural Experiment Station*

ACKNOWLEDGMENT OF PRIORITY

By John S. White

It has been called to my attention that the results in my note 'A t-test for the serial correlation coefficient' (Ann. Math. Stat., Dec. 1957) duplicate results obtained by M. H. Quenouille in 'Approximate tests of correlation in times-series 3' (Proc. Cambridge Phil. Soc., Vol. 45, part 3, 1949). I wish to acknowledge the priority of Prof. Quenouille's results which were inadvertently overlooked.

CORRECTION TO "ON THE POWER OF CERTAIN TESTS FOR INDEPENDENCE IN BIVARIATE POPULATIONS"

By H. S. Konijn

p. 304, line 13: like the left-hand side, the right-hand side is a function of n^*.
p. 305: beginning with the word "exists" Theorem 1.2 should read the same as Theorem 1.1, except that the exponent changes from $1/h$ to $1/hp^*$.
p. 306, line 1: change "of" to "at".
p. 309, line 3: insert "if ρ exists," preceding the expression for ER_n.
p. 309, last line of section 1: for $ER_n = 0$ read $ER_n \to 0$.
p. 309, line 8 of section 2: change "consist merely of" to "contain", and "or" to "plus".
p. 309, line 3 from below: change Λ to $\Lambda - \{\lambda^0\}$.

p. 310, line 1: change "is independent" to "is the distribution of two independent random variables".

CORRECTION TO "THE WAGR SEQUENTIAL T-TEST REACHES A DECISION WITH PROBABILITY ONE"

By Herbert T. David and William H. Kruskal

Two corrections to the paper of the above title (*Ann. Math. Stat.* Vol. 27 (1956), pp. 797–805) should be made.

(1) Page 803, line after (4.2): $K\sqrt{1 + K^2}$ should be replaced by $K/\sqrt{1 + K^2}$.
(2) Page 804, line 4: $v_n(A_n - R_n)$ should be replaced by $\sqrt{n}\ (A_n - R_n)$.

ABSTRACTS OF PAPERS

(Abstracts of papers presented for the Ames, Iowa Meeting of the Institute, April 3–5, 1958.)

41. Similar Tests of Hypotheses Concerning the Ratio of Mean to Standard Deviation in a Normal Population. Robert A. Wijsman, University of Illinois.

Let $X_1, \cdots, X_N$ be independent $N(\mu, \sigma^2)$ variables, and consider the hypothesis that μ/σ equals a given value against various alternatives. Let

$$T_1 = \sum X_i^2,\ T_2 = \sqrt{N}\bar{X}, \qquad T = (T_1, T_2),\ r = \sqrt{N}\mu/\sigma.$$

Then the density of T is $c(\sigma, r)h(t) \exp[-(t_1/2\sigma^2) + (r/\sigma)t_2]$ with $h(t) = (t_1 - t_2^2)^{n/2-1}$ if $t_1 \geqq t_2^2$ and $h = 0$ otherwise (we have put $n = N - 1$). Let the hypothesis be $r = r_0$. Associated with the exponential is a differential operator $D = \partial^2/\partial t_2^2 - 2r_0^2\partial/\partial t_1$. For a certain class C of functions G of t the test function $\alpha + \phi(t)$ with $\phi = h^{-1}DG$ will be similar and of size α. Conversely, to any similar test function $\alpha + \phi(t)$ there corresponds a $G \varepsilon C$, obtained by considering the differential equation $DG = h\phi$ as a heat (or diffusion) problem in one dimension, with a heat source density $h\phi$ which is a function of both position (t_2) and time (t_1), and solving the equation with help of the usual Green's function for the heat equation. Some of the unsolved problems concerning the search for an optimum similar test are indicated. (Rec. April 3, 1958)

(Abstracts of papers presented for the Los Angeles Meeting of the Institute, December 27–28, 1957.)

42. Demand for and Allocation of Engineering Personnel. I. Estimation of the Demand for Engineering Personnel, and General Formulation of the Allocation Problem. Rajendra Kashyap

Historical data for manpower and costs are analyzed for several types of contracts (prototype, initial, and follow-on contracts) with special regard to routines for (1) dis-

section of multiphase distributions with overlapping significant phases; (2) determination of standard patterns for incremental and cumulative manpower and costs; (3) estimation of total manpower and costs As to (1), graphical procedures may be useful (*Gibrat*, *Daeves*, etc.). For (2), the *Pearson* curve types may be applied, or the *Edgeworth-Kapteyn* system, which is closely related to the application of Hermitian polynomials, a method that for several reasons may deserve preference above all competing devices. (3) is a typical regression problem, the affinity and the effectivity of the chosen approach to be checked by *Fisher's* and *Student's* tests respectively. The problem of allocation of engineering personnel involves the determination of an optimal scheme for the allocation of available personnel to meet the demand for these personnel by the engineering units This allocation has to be satisfactory under surplus as well as under shortage conditions The simple consideration of manpower transfer to alternative fields of engineering activities shows clearly that optimization is necessarily an overall group problem. It can be described by an objective function considering competitive ability ratings in various fields, under the aspect of some suitable optimality criterion concerning costs, output or parametric quality-level Thus the complex problem is formally reduced to one in linear programming (Received March 14, 1958.)

43 Demand for and Allocation of Engineering Personnel. II. Integral-Valued Solutions of Allocation Problems. HERMAN W. VON GUERARD

Analysis of proportional representation, allocation or elimination of units is bound to integral-valued solutions. In consequence, proportionality, in general, can be approached only, and that leads to a problem of optimization. Unfortunately, that does not provide by itself the criterion for the least deviation from proportionality. Rounding procedures, in general, are not satisfactory. The main issue is, in terms of political elections, that no party is presumed to score less by the only reason that the total number of seats has been increased (postulate of monotony). Other criteria, based on least squares or on minimizing *Gram's* determinant (i.e maximizing linear dependence), are subject to the same considerations. The best expedient may be seen in requiring maximum likelihood to straight proportionality, and that is equivalent to sampling with replacement (the homogeneous case). The still more important procedure of sampling without replacement leads to *d'Hondt's* scheme (the inhomogeneous case), which is equivalent to maximum likelihood after adding one unit to each of the initial frequencies, i e to the popular votes per party Most of the related theorems can be easily visualized by multidimensional geometry of numbers (*Minkowski*), where *d'Hondt's* method of successive divisions is represented by successive penetrations of a vector through hyperplanes (Received March 14, 1958)

(Abstracts of papers presented for the Cambridge, Massachusetts Meeting of the Institute, August 25–30, 1958.)

44 On the Asymptotic Minimax Character of the Sample d.f. of Vector Chance Variables. J. KIEFER AND J. WOLFOWITZ, Cornell University. (By title)

Let $\mathfrak{F}$ (resp., $\mathfrak{F}^c$) denote the class of all d.f.'s (resp., continuous d.f.'s) on Euclidean m-space R^m. Let $X_1, \cdots, X_n$ be independent chance m-vectors with common unknown d f F The space D of decisions (values of the estimate of F) is any space of real functions d on R^m which includes all possible realizations of the sample d.f. S_n of $X_1, \cdots, X_n$. Let ϕ_n^* be the decision function which always makes decision S_n. Dvoretzky, Kiefer and Wolfowitz showed in *Ann. Math. Stat.*, 1956, that, when $m = 1$, ϕ_n^* is asymptotically minimax (as $n \to \infty$) for estimating F in $\mathfrak{F}$ or $\mathfrak{F}^c$, for any of a wide class of loss functions. In the present paper analogous results are proved when $m > 1$, despite the fact that S_n no longer

has the distribution-free property it has when $m = 1$. The resulting nonconstancy of the risk function $r(F, \phi_n^*)$ for F in $\mathfrak{F}^c$ and even the simplest loss functions, presents new difficulties in the minimax proof when $m > 1$: for example, the method of proof necessitates showing that $r(F, \phi_n^*)$ approaches a limit as $n \to \infty$, *uniformly* for F in an appropriately dense subset of $\mathfrak{F}^c$; the authors' results in *Trans. Amer. Math. Soc.*, 1958, are used in proving this. (Received March 21, 1958.)

45. Optimum Designs in Regression Problems. J. Kiefer and J. Wolfowitz, Cornell University. (By title)

Suppose Y_{x_i}, $i = 1, \cdots, n$, are independent random variables with $EY_x = \sum_1^k a_j f_j(x)$ for $x \varepsilon \mathfrak{X}$, where the f_j are known and the a_j are the unknown regression coefficients; Var $(Y_x) = v(x)\sigma^2$, where v is known. We consider the optimum allocation of the x_i for problems of statistical inference (1) about a_k, (2) about the s parameters $a_{k-s+1}, \cdots, a_k$, (3) about the whole function $\sum a_j f_j$. Algorithms are obtained which facilitate the computation of optimum designs (for several different optimality criteria, in the case of (2)). Examples are given which show the great simplification to be achieved by the use of these algorithms, over a more direct approach. For example, in case (1) the problem is solved by finding the best Chebyshev approximation to f_k of the form $\sum_1^{k-1} c_j f_j$ and locating the x_j, with appropriate frequencies, at points of maximum absolute deviation of the best approximation from f_k; in the example $\mathfrak{X} = [-1, 1], f_j(x) = x^{j-1}, k = h + 1$, the optimum design locates a fraction $1/h$ of the observations at each of -1 and 1 and a fraction $1/2h$ of the observations at $\cos (j\pi/h)$, $1 \leqq j \leqq h - 1$, and, as h increases, the relative efficiency of the often used "equal spacing" designs tends rapidly to zero. (Received April 17, 1958.)

46. Uniqueness of the L_2 Association Scheme. S. S. Shrikhande, University of North Carolina.

A partially balanced incomplete block design with $v = s^2$ treatments is said to have L_2 association scheme (R. C. Bose and T. Shimamoto, *Journal of the American Statistical Association*, 47: 151–184, 1952), if the treatments can be arranged in an $s \times s$ square such that any two treatments in the same row or the same column are 1-associates, whereas all the other pairs are 2-associates. In this case it is easily seen that $n_1 = 2s - 2$, $n_2 = (s - 1)^2$, $p_{11}^1 = s - 1$, $p_{11}^2 = 2$, where the symbols have the usual meanings. It is now proved that for a P.B.I.B. with s^2 treatments with the above values for n_1, n_2, p_{11}^1 and p_{11}^2, the association scheme is of L_2 type for all $s \geqq 3$ excepting $s = 4$. It can be shown that a necessary condition for existence of a symmetrical P.B.I.B. with above parameters, when s is even, is that $r - 2\lambda_1 + \lambda_2$ must be a perfect square and further $(r - \lambda_1 + (s - 1)(\lambda_1 - \lambda_2), -1) = 1$ for every odd prime p, where the last symbol stands for the Hilbert norm-residue symbol. The result contained in the last sentence, can also be obtained from a paper submitted by M. N. Vartak to the *Annals of Mathematical Statistics*. Here r, λ_1, λ_2 have the usual meaning. (Received May 26, 1958.)

47. On the existence of Wald's sequential test. Robert A. Wijsman, University of Illinois.

In the literature on Wald's sequential probability ratio test the question of existence of stopping bounds, given the two error probabilities, has never been answered. Granted existence, the uniqueness has been shown by L. Weiss (*Ann. Math. Stat.* Vol. 27 (1956) pp. 1178–1181) in the case that the probability ratio is continuous. Let α_1, α_2, be the two error

probabilities, and let $\alpha = (\alpha_1, \alpha_2)$. In the case of continuous probability ratio, and in the discrete case with suitable randomization, α_1 and α_2 are continuous functions of the stopping bounds Let C be the non-increasing (and convex) curve of points α produced by coincident stopping bounds, and let A be the set in the α-plane bounded by C and the coordinate axes Consider a point (α_1^*, α_2^*) on C, and separate the stopping bounds in a way which keeps α_1 constant. Since α_2 is a continuous function of the separation d between the bounds, with $\alpha_2(0) = \alpha_2^*$, $\alpha_2(\infty) = 0$, every value α_2 between 0 and α_2^* is assumed for some d. It follows that for every α in A there exist stopping bounds In the continuous case it is known from *Weiss' work that α_2 decreases monotonically from α_2^* to 0, as d increases from 0 to ∞* In that case, for the existence of stopping bounds it is also necessary that $\alpha \epsilon A$. (Received August 16, 1957; revised June 16, 1958)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of The Institute news items of interest

Personal Items

Gertrude Mary Cox, director of the Institute of Statistics, Consolidated University of North Carolina, was awarded an honorary Doctor of Science degree by Iowa State College during its Founder's Day centennial observance; she was cited as "teacher, researcher, leader and administrator in the field of statistics."

George Waddel Snedecor, who was primarily responsible for the development of the Iowa State College Statistical Laboratory, was awarded an honorary Doctor of Science degree by the college during its Founder's Day centennial observance and cited as "teacher, author, pioneer in experimental statistics." He has been a visiting professor at North Carolina State College, in the Institute of Statistics, since 1957.

Allan G. Anderson has resigned his position as Chief Statistician at the General Tire & Rubber Company, Akron, Ohio, to accept a position as Professor and Head of the Department of Mathematics at Western Kentucky State College, Bowling Green, Kentucky.

Dr. Ernst P. Billeter has been appointed Professor of Statistics and Automation at the University of Fribourg (Switzerland) He has also been elected Director of the Institute for Research in Automation, which has recently been founded at this University. The aim of this Institute is to do basic research work in application of automation in business and to introduce businessmen and their staff members, as well as students in economics, into the general methods of programming electronic data processing machines. Furthermore, this Institute will help businessmen in solving their problems in operations research, market research, and statistical quality control.

Dr. Uttam Chand has a new position as Officer on Special Duty (Training) in the Central Statistical Organisation (Cabinet Sectt.), New Delhi, India.

Dr. Frank A. Haight, formerly of Auckland University College, New Zealand, has returned to the United States to become Associate Mathematician at the Institute of Transportation and Traffic Engineering, U. C. L. A.

W. Robert Hydeman has accepted an appointment as Manager of Computer Systems at Touche, Niven, Bailey & Smart in their Executive Offices located at 1292 National Bank Building, Detroit 26, Michigan.

Richard C. Kao, formerly Research Associate, Operations Research Department, Engineering Research Institute, and Lecturer, Department of Mathematics, University of Michigan, Ann Arbor, is now Associate Mathematician, System Development Corporation, Santa Monica, California.

Mr. Frederick G. King recently took a position as Senior Scientist with the Armour Research Foundation and now lives in Evanston, Illinois. He was formerly with the Ballistic Research Laboratories at Aberdeen Proving Ground, Maryland.

1/Lt. Melville R. Klauber is now stationed with the 341st Air Refueling Squadron, Dow Air Force Base, Maine.

Richard A. Lamm, formerly at the Biological Warfare Laboratories, Fort Detrick, Maryland, is now a Statistician with the American Cyanamid Company at Pearl River, New York.

Dr. William G. Madow has been advanced to the position of Staff Scientist of Stanford Research Institute, Menlo Park, California.

William F. Taylor has left the School of Aviation Medicine, Randolph Air Force Base, Texas, to become Associate Professor of Public Health in the Division of Biostatistics of the School of Public Health, University of California at Berkeley.

H. Robert van der Vaart, who has been a visiting professor at the Department of Experimental Statistics of the Institute of Statistics at Raleigh, North Carolina, from January, 1957, until the end of January, 1958, will be a visiting associate professor (and hold a scholarship from the Netherlands Organization for Pure Research, Z.W.O.) at the Department of Statistics of the University of Chicago.

Ronald E. Walpole has completed the requirements for the Ph.D. degree in Statistics at Virginia Polytechnic Institute and has assumed the position of Head of the Department of Mathematics and Statistics at Roanoke College.

New Members

The following persons have been elected to membership in The Institute

Febr… 958, to May 13, 1958

Agan, Miss Martha L., B.S. (Uni… rnia, Los Ang… Medical Record Librarian, V. A. Center, Los Angeles, … *1814 Holmby … Los Angeles 25, California.*

Alling, David W., M.D. (Univ. …), Student, … versity, Ithaca, New York; *1124 Ellis Hollow Roa… York.*

… **Laurence H.**, B.S. (Iowa … , Research … epartment of Animal *bandry, University of Mi… ul, Minnes…*

… **George A.**, M.A. (Ca… ssor of M… Statistics, *Imper… Mathematics Depart… of Londo… Road, London*

Berliner, Paul, M.B.A. (City College of New York), Engineer, *Radio Corporation of America, Depart 660, 18-3, 415 South 5th St., Harrison, New Jersey.*

Blair, Charles R., B S. (George Washington Univ.), Mathematician, National Security Agency, Ft. George G. Meade, Maryland, and student, George Washington Univ., Washington, D. C.; *536 Beacon Road, Silver Spring, Maryland.*

Cohen, F. A., M.A. (U C.L.A.), Teaching Assistant, University of California at Los Angeles, Los Angeles 24, California; *11651 Gorham Ave., #4, Los Angeles 49, California*

Cunia, Tiberius, M S. (McGill Univ), Forest Engineer, Canadian International Paper Co., 1461 Sunlife Building, Montreal, Quebec, Canada; *5562 Basile Patenaude Pl , Montreal, Quebec, Canada*

Dutt, John E., M A (Columbia Univ.), Mathematician, MIT Lincoln Laboratory, Lexington, Massachusetts; *55 Arlington Street, Newton, Massachusetts*

Elashoff, Robert M., A.M. (Boston Univ.), Student and Laboratory Teacher in Biostatistics, *Harvard School of Public Health, 55 Shattuck Street, Boston, Massachusetts*

Ferrin, Kenneth M., M.A. (U C.L A.), student, U C L A ; *1412 Midvale Avenue, West Los Angeles 24, California.*

Federowicz, Alexander J., B S (Carnegie Inst of Tech.), Graduate Student, Carnegie Institute of Technology, Pittsburgh 13, Pa ; *5876 Solway Street, Pittsburgh 17, Pa*

Fimple, Melvin D., M.B.A. (Univ. of Buffalo), Components Engineer, Stromberg-Carlson Company, Rochester, New York; *21 Carthage Drive, Rochester 21, New York*

Fink, Lester H., B.S in E.E. (Univ. of Pennsylvania), Engineer, Electrical Research Section, Philadelphia Electric Co., Philadelphia, Pa.; *Ferry and Iron Hill Roads, Doyletown R. D. 1, Pa*

Freimer, Marshall Leonard, A M. (Harvard Univ.), Student, Harvard University, Dept of Statistics, Cambridge 38, Massachusetts, *Lincoln Laboratory, P O. Box 73, Lexington 73, Massachusetts.*

Grossling, Bernardo F., Ph.D. (London Univ.), Senior Research Geophysicist, *California Research Corporation, P. O. Box 446, La Habra, California.*

Howell, John Robert, M S. (Univ. of Florida), graduate student, University of Florida, Gainesville, Florida; *Mathematics Department, University of Florida, Gainesville, Florida*

Johnson, Jerome R., M.S. (Purdue Univ), Chief, Rocket, Mortar & Recoilless Ammunition Section, Surveillance Branch, Weapon Systems Lab., Ballistic Research Lab , Aberdeen Proving Ground, Maryland; *860 Ontario Street, Havre de Grace, Maryland*

Kaula, William M., M.S. (Ohio State Univ.), Geodesist, U S Army Map Service, Washington 25, D. C.; *5202 Baltimore Avenue, Washington 16, D. C*

Lerner, Gary B., B S. (Michigan State Univ.), Actuarial Student, Metropolitan Life Insurance Company, 1 Madison Ave., New York, N Y.; *731 Scranton Ave., East Rockaway, New York.*

Lewis, John S., B.S. (Carnegie Inst. of Tech.), Research Assistant, *Department of Mathematics, Carnegie Institute of Technology, Pittsburgh 13, Pennsylvania.*

Meagher, Jack R., A.M. (Univ. of Michigan), Associate Professor, *Mathematics Department, Western Michigan University, Kalamazoo, Michigan.*

Posener, Ludwig N., Ph.D. (Univ. of Berlin), Lecturer of Statistics and Applied Mathematics, University of Tel Aviv, 155 Herzl Street, Tel Aviv, Israel; *22 Pinsker Street, Rehovot, Israel.*

Raj, Des, Ph D. (Calcutta Univ.), *Associate Professor, American University of Beirut, Beirut, Lebanon.*

Sawits, Murray B., B.S. (City College of New York), Senior Statistician, Rayco Mfg Co , 22 Straight St., Paterson, New Jersey; *1420 Grand Concourse, Box 56, New York, N. Y.*

Suzuki, Yukio, Member of *the Institute of Statistical Mathematics, No. 1, Azabu-Fujimi-Cho, Minato-Ku, Tokyo, Japan*

Thompson, Robert J., B.S (Drake Univ.), Senior Research Engineer, Convair Pomona, Pomona, California; *447 Celia Avenue, Pomona, Calif.*

Zadoff, Solomon A., A.M. (Columbia Univ.), Research Engineer, Sperry Gyroscope Co ,

Great Neck, New York, and student, Columbia University, New York, New York; *193-18 37th Ave., Flushing 69, New York.*

Zayachkowski, Walter, M.A. (Univ. of Saskatchewan), Graduate Student, *Dept. of Mathematics, University of Alberta, Edmonton, Alberta, Canada.*

Zimmer, William J., M.S. (Purdue Univ.), Research Fellow, *Statistical Laboratory, Purdue University, Lafayette, Indiana.*

EXPANDED TRAINING PROGRAM IN BIOMETRICS TO BE OFFERED AT IOWA STATE COLLEGE STATISTICAL CENTER

The Department of Statistics and the Statistical Laboratory of Iowa State College will substantially expand their present graduate training program in biostatistics with the aid of a five-year grant from the National Institutes of Health. This award will provide support for several graduate students in statistics per year as candidates for the M.S. or Ph.D. degree, with a view to stimulating their interest in biometry, medical statistics or public health as a career. It will also give partial support to one staff member so that he can devote more time to those areas of statistical application.

One feature of the expanded program is that biostatistics trainees, while working toward masters' or doctors' degrees in statistics, will spend up to three months each year at some selected medical school or public health center to round out their experience through contact with biometric data in the field or laboratory. So far, three new traineeships have been established for the 1958–59 year. Further details about the expanded biostatistics program and application forms for traineeships for the 1959–60 year may be obtained from the Department of Statistics, Iowa State College.

NATIONAL REGISTER OF SCIENTIFIC AND TECHNICAL PERSONNEL

The American Mathematical Society at the request of the National Science Foundation is assembling and maintaining a register of mathematicians and mathematical scientists. The Mathematics Register is a section of the National Register of Scientific and Technical Personnel, which is an official responsibility of the NSF. The purpose of the Register is to provide up-to-date information on the scientific manpower resources of the United States.

As a result of the splendid cooperation accorded to the project by most of the mathematicians and mathematical scientists who have received questionnaires to fill in, the mathematical section of the Register is now remarkably complete. However, there are still a few gaps to be filled in. If you have received a National Register questionnaire from the Society, please fill it in now and send it to the Headquarters Offices of the Society at 190 Hope Street, Providence 6, Rhode Island. If you have never received a questionnaire and feel that you are qualified for inclusion in the Register, please drop a note to that effect to the Society at this address.

EDUCATIONAL TESTING SERVICE FELLOWSHIPS

The Educational Testing Service is offering for 1959–60 its twelfth series of research fellowships in psychometrics leading to the Ph.D. degree at Princeton University. Open to men who are acceptable to the Graduate School of the University, the two fellowships each carry a stipend of $2,650 a year and are normally renewable. Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School.

Suitable undergraduate preparation may consist either of a major in psychology with supporting work in mathematics, or a major in mathematics together with some work in psychology. However, in choosing fellows, primary emphasis is given to superior scholastic attainment and research interests rather than to specific course preparation.

The closing date for completing applications is January 2, 1959. Information and application blanks will be available about September 15 and may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

REPORT OF THE AMES, IOWA, MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The seventy-sixth meeting of The Institute of Mathematical Statistics, a Central Region Meeting, was held in the Gallery of the Memorial Union on the campus of Iowa State College at Ames, Iowa, on April 3–5, 1958. These dates were within the period during which Iowa State College was observing its Centennial Celebration.

A Special Invited Address, "Subjective Judgements and Statistical Practice," was delivered by Professor L. J. Savage of the University of Chicago.

On Friday evening, April 4, a banquet was held in the Great Hall of the Memorial Union with Professor T. A. Bancroft presiding. After dinner Dean Richard S. Bear of the Division of Science at Iowa State College addressed the assembled guests on the history of statistics at Ames. This was followed by entertainment by graduate students at Ames.

The Chairman of the Program Committee for the meeting was Jack Silber, Roosevelt University. The Assistant Secretary for the meeting was Herbert T. David, Iowa State College.

Ninety-six people registered for the meetings, including the following 52 members of The Institute:

D Huntsberger, Meyer Dwass, Preston C. Hammer, Emil H. Jebe, Howard L Taylor, H. O Hartley, W. M. Gilbert, Paul G. Homeyer, Franklin A. Graybill, W. H Horton, Oscar Kempthorne, Russell N. Bradt, D. R. Truax, Stanley Isaacson, I. R. Savage, Lorraine Schwartz, George Zyskind, Helen Bozivich, Robert V. Hogg, Leo Katz, Leonard J Savage, Roger S. McCullough, Howard L. Jones, Scott Krane, F. E. Satterthwaite, Bernard Ostle, Virgil S Anderson, R. W. Kennard, Herbert T. David, Robert F. White, A. W.

ham, J. D. Hromi, Jack Silber, John F. Pauls, Timon A. Walther, Edward C. Bryant, ard L. Beatty, Richard L. Carter, Byron Brown, S. N. Roy, Betty K. Stewart, Z. ndarajulu, M. B. Wilk, H. Robert van der Vaart, William H. Williams, T. A. Ban-, William J. Zimmer, R. A. Wijsman, G. Tintner, John Gurland, Sidney Addelman, d L. Wallace.

ıe program for the meeting was as follows:

THURSDAY, APRIL 3, 1958

0 a.m. Invited Papers on the Design of Experiments

:man: VIRGIL L. ANDERSON, Purdue University

1. *A Comparison of Designs for Exploration of Response Surfaces*, LE ROY FOLKS, Iowa State College.
2. *The Staircase Design*, F. A. GRAYBILL, Oklahoma State University.

0 a.m. Invited Papers on the Problem of Nuisance Parameters

:man: T. A. BANCROFT, Iowa State College.

1. *Testing the Equality of the Means of Two Normal Populations*, JOHN GURLAND, Iowa State College.
2. *The Behrens—Fisher Problem: A Critical Review and a Subjective Approach*, DAVID L. WALLACE, The University of Chicago.

0 p.m. Special Invited Address

'man: OSCAR KEMPTHORNE, Iowa State College

Subjective Judgements and Statistical Practice, L. J. SAVAGE, The University of Chicago.

0 p.m. Invited Paper on the Analysis of Variance

'man: R. V. HOGG, University of Iowa.

1. *Multivariate Analysis of Variance under Models I and II and Mixed Models*, S. N. ROY, University of North Carolina and University of Minnesota.

FRIDAY, APRIL 4, 1958

) a.m. Invited Papers on Statistical Problems in Econometric Theory

'man: J. SILBER, Roosevelt University.

1. *A New Method for Fitting the Logistic Function*, GERHARD TINTNER, Iowa State College.
2. *The Effects of Incomplete Specification on the Results of Estimating Procedures*, LEONID HURWICZ, University of Minnesota.

) a.m. Contributed Papers I

man: ALBERT WORTHAM, Texas Instruments.

1. *Bias and Confidence in Not-Quite Large Samples* (Preliminary Report), JOHN W. TUKEY, Princeton University (By title).
2. *On a Multivariate Gamma Distribution*, P. R. KRISHNAIAH and M. M. RAO, University of Minnesota.
3. *On the Fitting of Some Contagious Distributions*, S. K. KATTI and JOHN GURLAND, Iowa State College.
4. *Minimal Complete Classes of Tests*, D. L. BURKHOLDER, University of Illinois

5 *An Identity of Use in Non-Linear Least Squares*, M. B. WILK, Bell Telephone Laboratories.
6 *Contributions to the Theory of Rank Order Statistics—The One Sample Case*, I. RICHARD SAVAGE, University of Minnesota.
7. *A Rule for Action Based on Percentage Changes in the Sample Mean*, D. B OWEN, Sandia Corporation (By title).
8. *An Expression for the Cumulative Distribution Function of the Non-Central t-Distribution*, D. B OWEN, Sandia Corporation (By title).
9 *Some Formulae for the Exact Computation of Probabilities in Wilcoxon's Two Sample Test*, H ROBERT VAN DER VAART, University of Chicago

2:00 p.m. Invited Papers on Non-Parametric Statistics

Chairman B. OSTLE, Sandia Corporation.
1. *Some Null Rank Distributions Derivable by Reflection*, H. T DAVID, Iowa State College.
2 *Order Statistics in the Poisson Process*, MEYER DWASS, Northwestern University

3:30 p.m. Invited Papers on the Use of Electronic Computers in Statistics

Chairman M B WILK, Bell Telephone Laboratories
1. *Theoretical Possibilities of Computers*, P C HAMMER, University of Wisconsin
2. *Linear Programming on the IBM-650*, H O HARTLEY, Iowa State College

SATURDAY, APRIL 5, 1958

9:00 a.m. Contributed Papers II

Chairman · W. H. HORTON, Westinghouse Electric Company.
1 *Biases in Prediction by Regression for Certain Incompletely Specified Models*, HAROLD LARSON, Iowa State College
2 *Notes on the Spearman-Karber Procedures in Bioassay* (Preliminary Report), BYRON W. BROWN, JR , Louisiana State University.
3. *Approximate Solutions for the Probability Density of Zero-Crossing Intervals in a Gaussian Process*, J. A. McFADDEN, Naval Ordnance Laboratory and Purdue University (introduced by JUDAH ROSENBLATT)
4 *The Fourth Product-Moment of a Binary Random Process*, J. A McFADDEN, Purdue University (introduced by JUDAH ROSENBLATT). (By title)
5. *Limiting Distributions of k-Sample Test Criteria of Kolmogorov-Smirnov-v Mises Type*, J. KIEFER, Cornell University. (By title)
6 *Independence of Statistics and Characterization of the Multivariate Normal Distribution*, S. G. GHURYE, University of Chicago, and INGRAM OLKIN, Michigan State University.
7. *Unbiased Regression Estimators*, WILLIAM H. WILLIAMS, Iowa State College.
8 *Maximum Likelihood Estimation from Incomplete Data for Continuous Distribution*, SCOTT KRANE, Iowa State College.
9. *Unbiased Ratio Estimators in Stratified Sampling*, JOSE NIETO, Iowa State College.
10 *Similar Tests of Hypothesis Concerning the Ratio of Mean to Standard Deviation in a Normal Population*, ROBERT A. WIJSMAN, University of Illinois.
11 *Births and Deaths in Parallel*, J. SILBER, Roosevelt University.

10:30 a.m. Invited Papers on the Theory of Estimation

Chairman · R N. BRADT, Uuniversity of Kansas.
1 *Some Interval Estimation Problems*, ROBERT J. BUEHLER, Iowa State College

2. *Inadmissible Samples and Confidence Limits*, HOWARD L. JONES, Illinois Bell Telephone Company.

On Saturday, April 5, at 2:00 p.m. Dr. S. N. Roy of the University of North Carolina and the University of Minnesota presented a special seminar for members of the Statistical Laboratory at Iowa State College on "Some Recent Work on Univariate and Multivariate Components Analysis." The people attending the meetings of The Institute were invited to this seminar.

PUBLICATIONS RECEIVED

ARROW, KENNETH J., SAMUEL KARLIN, AND HERBERT SCARF, *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, California. x + 340 pp.

A Comparative Study of Statistical Analysis and Other Methods of Computing Ore Reserves, United States Department of the Interior Bureau of Mines, Washington 25, D. C.

Supplementary List of Publications of the National Bureau of Standards, July 1, 1947, to June 30, 1957. Supplement to National Bureau of Standards Circular 460. (Supersedes Supplement to Circular 460, December 30, 1952.) Issued May 14, 1958, 373 pages, $1.50. (Order from Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.)

ON MIXED SINGLE SAMPLE EXPERIMENTS[1]

By Leonard Cohen

College of the City of New York

1. Introduction and summary. William Kruskal [1], Howard Raiffa [2], J. L. Hodges, Jr. and E. L. Lehmann [4], have shown that in certain Neyman-Pearson type problems of testing a simple hypothesis against a simple alternative, determining the sample size by means of a chance device yields improvements over fixed sample size procedures. The purpose of this paper is not only to investigate the general problem of randomizing over fixed sample size tests of a simple hypothesis against a simple alternative, but also randomizing over other fixed sample size procedures in topics such as confidence interval estimation, the k-decision problem, etc.

In Section 2, a fixed sample size test of a simple hypothesis against a simple alternative is identified with an operating characteristic (α, β, n) where α denotes the probability of a type I error, β denotes the probability of a type II error, and n denotes the sample size. A mixed single sample test is defined as a sequence of quadruples.

$(\gamma_i, \alpha_i, \beta_i, n_i)$, where $\gamma_i \geqq 0, \sum_{i=1}^{\infty} \gamma_i = 1$, where (α_i, β_i, n_i) is a fixed sample size test and where γ_i is interpreted as the probability of using the fixed sample size test (α_i, β_i, n_i) for $i = 1, 2, \cdots$. A mixed single sample test is identified with an operating characteristic $(\alpha, \beta, n) = \sum_{i=1}^{\infty} \gamma_i(\alpha_i, \beta_i, n_i)$. For each non-negative integer n, the class A_n of admissible fixed sample size procedures of sample size n is defined in an obvious way. We define $A = \bigcup_{i=0}^{\infty} A_i$ and A^* as the convex hull of A. It is not necessarily true that A^* is closed. An example is given to show this. However, it is true that the lower boundary of A^* is a subset of A^* so that the lower boundary of A^* determines a minimally complete class, $\mathcal{A}$, of mixed single sample tests. The tests in $\mathcal{A}$ are characterized from a Bayes point of view and a technique for constructing the tests in $\mathcal{A}$ is given.

In Section 3, the technique is applied to tests on the mean of a normal distribution with known variance. It is shown that the tests in $\mathcal{A}$ are either

(a) fixed sample size tests, or

(b) mixtures of at most two fixed sample size tests.

It is shown that there exists a minimal subset $\mathcal{A}_0$ of A such that all improved randomized procedures are of the form $(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$ or $(\alpha, \beta, n) = \gamma(1, 0, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$, where $0 < \gamma < 1$ and where $(\alpha_0, \beta_0, n_0) \in \mathcal{A}_0$. It is then shown how to construct $\mathcal{A}_0$. The following problems (of the Neyman-Pearson type) are solved:

(a) Given α and β, how can we find the test in $\mathcal{A}$ with the given α and β?

Received January 2, 1958.

[1] Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Faculty of Political Science, of Columbia University, 1957.

(b) Given α and n, how can we find the test in $\mathcal{A}$ with the given α and n? Numerical examples are worked out.

In Section 4, the technique is applied to tests on the mean of a binomial distribution. Although no general results were obtained, numerical examples of interest are given.

In Section 5, the technique is applied to tests on the range of a rectangular distribution (when one end point is known). It is shown that if $\alpha > 0$, $n > 0$, and $(\alpha, \beta, n) \varepsilon A_n$, then $(\alpha, \beta, n) \not\varepsilon \mathcal{A}$. The tests in $\mathcal{A}$ are characterized by a simple equation which makes it easy to

(a) determine whether a given point (α, β, n) belongs to $\mathcal{A}$, and

(b) construct any test in $\mathcal{A}$, given two of the three coordinates.

It is shown that if $(\alpha, \beta, n) \varepsilon A_n$, then there exists a test (α, β, n') in $\mathcal{A}$ such that $n' = (1 - \alpha)n$. Hence, the fractional saving in the expected sample size achieved by randomization is equal to α.

In Section 6, it is shown that in tests on the mean of a rectangular distribution (with known range), it never pays to randomize.

In Section 7, confidence intervals are evaluated in terms of confidence coefficient (α), expected length (L) and expected sample size (n). For the problem of obtaining a confidence interval for the mean of a normal distribution with known variance, "improved" randomized procedures exist and are of the form $(\alpha, L, n) = \gamma(0, 0, 0) + (1 - \gamma)(\alpha', L', n')$ where $0 < \gamma < 1$ and where (α', L', n') is a fixed sample size confidence interval procedure. Clearly, the randomized procedures obtained are of such a nature that the question of confidence intervals evaluated in terms of expected length and/or expected sample size is thrown open to discussion.

In Section 8, the k-decision problem is discussed. It is shown that improvements can be obtained by randomization.

In Section 9, the problem of applying mixed single sample tests of a composite hypothesis against a composite alternative is discussed.

In Section 10, mixed single sample procedures are compared to Wald's sequential probability ratio test in the problem of tests on the range of a rectangular distribution when one endpoint is known and are shown to be efficient in a certain sense.

In Section 11, the estimation problem is mentioned. It is shown that in most practical problems, fixed sample size procedures are optimal.

In Section 12, applications of mixed single sample tests are discussed.

2. Testing a simple hypothesis against a simple alternative. Let X denote a random variable with density function (or discrete probability function) $f(x, \theta)$. We wish to test the hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$. In the sequel, we shall restrict ourselves exclusively to fixed sample size tests, both randomized and non-randomized, and mixtures of such tests. Any test of the preceding kinds will be identified with an operating characteristic (α, β, n), where α denotes the probability of a type I error, β denotes the probability of a

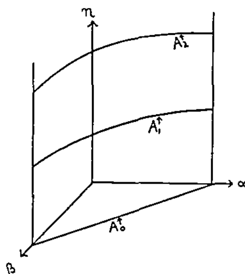


FIG. 1

type II error, and n denotes the expected number of observations. If two tests have the same operating characteristic, they will be considered equivalent.

Let $(x_1, x_2, \cdots, x_n)$ denote a sample of n independent observations on X. Let δ_n denote a real-valued, measurable function of n variables whose range is the closed interval (0, 1). The expression $\delta_n(x_1, x_2, \cdots, x_n)$ is interpreted as the probability of rejecting H_0 if $(x_1, x_2, \cdots, x_n)$ is observed. Let Δ_n denote the class of functions $\{\delta_n\}$ of the preceding type.

Definition 1. For any integer $n > 0$, let $S_n = \{(\alpha, \beta, n): \alpha =$ [illegible] $E(1 - \delta_n \mid \theta_1), \delta_n \varepsilon \Delta_n\}$. *$S_n$ is the class of tests of fixed sample size* [illegible] $S_0 = \{(\alpha, \beta, 0): 0 \leqq \alpha \leqq 1, \alpha + \beta = 1\}$

Definition 2. For any integer $n \geqq 0$, let $A_n = \{(\alpha, \beta, n): a)(\alpha,$ [illegible] there exists no other test (α', β', n) belonging to S_n with the property [illegible] $\beta' \leqq \beta$, at least one of these inequalities being strict.$\}$
The set A_n is the class of admissible procedure based on [illegible] and is known to be complete. See Fig. 1.

Definition 3. Let $A = \bigcup_{i=0}^{\infty} A_i$.

Definition 4. Let $A^* = \{(\alpha, \beta, n): (\alpha, \beta, n) = \sum_{i=0}^{\infty} \gamma_i$ [illegible] $\sum_{i=0}^{\infty} \gamma_i = 1$, and $(\alpha_i, \beta_i, n_i) \varepsilon A$ for $i = 0, 1, 2, \cdots \}$.
γ_i is interpreted as the probability of selecting [illegible] (α_i, β_i, n_i) for $i = 0, 1, 2, \cdots$. A^* is the convex [illegible]

Definition 5. Let $\mathfrak{A} = \{(\alpha, \beta, n): a)(\alpha, \beta, n) \varepsilon A^*,$ [illegible] test (α', β', n') belonging to A^* with the property [illegible] at least one of these inequalities being strict.$\}$.
The set $\mathfrak{A}$ is the class of admissible mixed single [illegible]

We next wish to show that $\mathfrak{A}$ is complete, i.e., [illegible] there exists a test (α, β, n) in $\mathfrak{A}$ such that $\alpha \leqq \alpha'$ [illegible] these inequalities being strict. If, in general, A^* [illegible] $\mathfrak{A}$ is complete. However, A^* is not necessarily [illegible] will illustrate.

Example. Let $f(x, \theta) = 1$ if $\theta \leqq X \leqq \theta + 1$

$= 0$ elsewhere.

We wish to test the hypothesis $H_0: \theta = 0$ against the alternative $H_1: \theta = \theta_1$, where $0 < \theta_1 < 1$. A simple calculation shows that

$$(1) \qquad A_n = \{(\alpha, \beta, n): 0 \leqq \alpha \leqq (1 - \theta_1)^n, \alpha + \beta = (1 - \theta_1)^n\}.$$

We define a sequence $\{(\alpha_k, \beta_k, n_k)\}$, where $(\alpha_k, \beta_k, n_k) = (1 - 1/k)(0, 1, 0) + (1/k)(0, (1 - \theta_1)^k, k)$. Clearly $\lim_{k\to\infty}(\alpha_k, \beta_k, n_k) = (0, 1, 1)$. However, $(0, 1, 1) \not\varepsilon A^*$. To prove this, assume $(0, 1, 1) \varepsilon A^*$. Then, since A^* is a three-dimensional convex set, $(0, 1, 1)$ can be expressed as a convex linear combination of at most four points in A, i.e., $(0, 1, 1) = \sum_{i=1}^4 \gamma_i(\alpha_i, \beta_i, n_i)$, where $\gamma_i \geqq 0, \sum_{i=1}^4 \gamma_i = 1$ and $(\alpha_i, \beta_i, n_i) \varepsilon A$ for $i = 1, 2, 3, 4$. Since $\sum_{i=1}^4 \gamma_i\beta_i = 1$, it follows that $\beta_i = 1$ if $\gamma_i > 0$. However, if $\beta_i = 1$, it follows from (1) that $n_i = 0$, contradicting the assumption $\sum_{i=1}^4 \gamma_i n_i = 1$. Q.E.D.

In order to show that $\mathfrak{A}$ is complete, we define $A_L^* = \{(\alpha, \beta, n)$: (a) (α, β, n) is a boundary point of A^*, and (b) there exists no test (α', β', n') belonging to A^* such that $\alpha' \leqq \alpha, \beta' \leqq \beta, n' \leqq n$, at least one of these inequalities being strict.$\}$.

The set A_L^* is the "lower" boundary of A^*. Clearly, $\mathfrak{A} \subset A_L^*$. We shall now prove an important theorem.

THEOREM 1. $A_L^* \subset A^*$.

PROOF. Suppose $(\alpha, \beta, n) \varepsilon A_L^*$. Then, since (α, β, n) is a boundary point of A^*, there exists a sequence of points $\{(\alpha_k, \beta_k, n_k)\}$ belonging to A^* such that

$$(\alpha, \beta, n) = \lim_{k\to\infty}(\alpha_k, \beta_k, n_k).$$

Since A^* is a three dimensional convex set, each point (α_k, β_k, n_k) of this sequence can be expressed as a convex linear combination of at most four points in A: i.e., for each k, there exist numbers $\gamma_{ik}, \alpha_{ik}, \beta_{ik}, n_{ik}$ such that $(\alpha_k, \beta_k, n_k) = \sum_{i=1}^4 \gamma_{ik}(\alpha_{ik}, \beta_{ik}, n_{ik})$, where $\gamma_{ik} \geqq 0, \sum_{i=1}^4 \gamma_{ik} = 1$ and $(\alpha_{ik}, \beta_{ik}, n_{ik}) \varepsilon A$ for $i = 1, 2, 3, 4$. Without any loss of generality, we can assume that the sequences $\{\gamma_{ik}\}$, $\{\alpha_{ik}\}$ and $\{\beta_{ik}\}$ are convergent for $i = 1, 2, 3, 4$ as k tends to infinity. Let $\gamma_i = \lim_{k\to\infty} \gamma_{ik}$, $\alpha_i = \lim_{k\to\infty} \alpha_{ik}$, $\beta_i = \lim_{k\to\infty} \beta_{ik}$ for $i = 1, 2, 3, 4$. Clearly, $\gamma_i \geqq 0, \sum_{i=1}^4 \gamma_i = 1$ for $i = 1, 2, 3, 4$. Before proceeding with the proof of Theorem 1, we prove a useful lemma.

LEMMA 1. *If $\gamma_i > 0$, there exists a number N_i such that $n_{ik} \leqq N_i$ for all k.*

PROOF. Since $\lim_{k\to\infty} n_k = n$, there exists a positive integer K such that if $k > K$, $n_k < n + 1$. Furthermore, since $\lim_{k\to\infty} \gamma_{ik} = \gamma_i > 0$, there exists a number K_i such that if $k > K_i$, $\gamma_{ik} > \frac{1}{2}\gamma_i$. Let $M_i = \max(K, K_i)$. Then, if $k > M_i$, $\frac{1}{2}\gamma_i n_{ik} \leqq \sum_{i=1}^4 \gamma_{ik} n_{ik} = n_k < n + 1$. Thus, if $k > M_i$, $n_{ik} < 2(n + 1)/\gamma_i$. Then, $N_i = \max[n_{i1}, n_{i2}, \cdots n_{iM_i}, 2(n + 1)/\gamma_i]$ is the required number, proving the lemma.

We now proceed with the proof of Theorem 1. Consider four cases.

Case 1. $\gamma_i > 0, i = 1, 2, 3, 4$.

Let $N = \max(N_1, N_2, N_3, N_4)$ where N_i is defined in Lemma 1. Then, since

$0 \leqq n_{ik} \leqq N$, for $i = 1, 2, 3, 4$ and all k, the sequences $\{n_{ik}\}$ are bounded. Hence, for each i, there exists a convergent subsequence which we denote by $\{\bar{n}_{ik}\}$. Let $\{\bar{\alpha}_{ik}\}$, $\{\bar{\beta}_{ik}\}$ and $\{\bar{\gamma}_{ik}\}$ denote respectively the subsequences of $\{\alpha_{ik}\}$, $\{\beta_{ik}\}$ and $\{\gamma_{ik}\}$ corresponding to the convergent subsequence $\{\bar{n}_{ik}\}$ of $\{n_{ik}\}$. Let $\lim_{k\to\infty} \bar{n}_{ik} = n_i$. Clearly,

$$(\alpha, \beta, n) = \lim_{k\to\infty} (\alpha_k, \beta_k, n_k) = \lim_{k\to\infty} \sum_{i=1}^{4} \gamma_{ik}(\alpha_{ik}, \beta_{ik}, n_{ik}) = \sum_{i=1}^{4} \lim_{k\to\infty} \gamma_{ik}(\alpha_{ik}, \beta_{ik}, n_{ik})$$

$$= \sum_{i=1}^{4} \lim_{k\to\infty} \bar{\gamma}_{ik}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) = \sum_{i=1}^{4} \gamma_i(\alpha_i, \beta_i, n_i).$$

Since $(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) \varepsilon A$ for all i and k, and since A is closed,

$$\lim_{k\to\infty}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) = (\alpha_i, \beta_i, n_i) \in A \qquad \textit{for } i = 1, 2, 3, 4.$$

Furthermore, since $\gamma_i > 0$, $i = 1, 2, 3, 4$, and $\sum_{i=1}^{4} \gamma_i = 1$, $\sum_{i=1}^{4} \gamma_i(\alpha_i, \beta_i, n_i) \varepsilon A^*$. Hence $(\alpha, \beta, n) \varepsilon A^*$.

The more difficult case to prove is Case 2

Case 2. Exactly one of the γ_i's is 0.

To fix ideas, suppose $\gamma_1 = 0$, $\gamma_2 > 0$, $\gamma_3 > 0$, $\gamma_4 > 0$. Let $N = \max(N_2, N_3, N_4)$ In a manner analogous to that used in Case 1, we define sequences $\{\bar{\alpha}_{ik}\}$, $\{\bar{\beta}_{ik}\}$, $\{\bar{n}_{ik}\}$ and $\{\bar{\gamma}_{ik}\}$ for $i = 2, 3, 4$. We define new sequences

$$\alpha_k' = \bar{\gamma}_{1k}(0) + \sum_{i=2}^{4} \bar{\gamma}_{ik}\, \bar{\alpha}_{ik},$$

$$\beta_k' = \bar{\gamma}_{1k}(1) + \sum_{i=2}^{4} \bar{\gamma}_{ik}\, \bar{\beta}_{ik},$$

$$n_k' = \bar{\gamma}_{1k}(0) + \sum_{i=2}^{4} \bar{\gamma}_{ik}\, \bar{n}_{ik},$$

where $\bar{\gamma}_{1k} = 1 - \sum_{i=2}^{4} \bar{\gamma}_{ik}$. It is easily seen that

$$\lim_{k\to\infty} \alpha_k' = \alpha,$$

$$\lim_{k\to\infty} \beta_k' = \beta,$$

$$\lim_{k\to\infty} n_k' \leqq n,$$

Since $(\alpha_k', \beta_k', n_k') \varepsilon A^*$ for each k, and since $(\alpha, \beta, n) \varepsilon A_L^*$, it follows that the inequality $\lim_{k\to\infty} n_k' < n$ cannot hold. Hence,

$$(\alpha, \beta, n) = \lim_{k\to\infty} (\alpha_k', \beta_k', n_k') = \lim_{k\to\infty} \sum_{i=2}^{4} \bar{\gamma}_{ik}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik})$$

$$\cdot \sum_{i=2}^{4} \lim \bar{\gamma}_{ik}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) = \sum_{i=2}^{4} \gamma_i(\alpha_i, \beta_i, n_i).$$

Using the argument in Case 1, we find that $(\alpha, \beta, n) \varepsilon A^*$.

Case 3. Two of the γ_i's are 0. The proof of Case 3 is analogous to the proof of Case 2.

Case 4. Three of the γ_i's are 0. The proof of Case 4 is analogous to the proof of Case 2.

COROLLARY 1. $\mathcal{A} = A_L^*$.

COROLLARY 2. *$\mathcal{A}$ is complete.*

PROOF. Let (α', β', n') be a test which does not belong to $\mathcal{A}$.

Let

$$A_{(\alpha'\beta')} = \{(\alpha, \beta, n){:}a)\ \alpha = \alpha', \beta = \beta' \text{ and } (\alpha, \beta, n) \in A^*\}.$$

$A_{(\alpha',\beta')}$ is non-empty since $(\alpha', \beta', n') \in A_{(\alpha',\beta')}$. Let

$$N = N_{(\alpha',\beta')} = \inf_{\{n':(\alpha',\beta',n') \in A_{(\alpha',\beta')}\}} n'$$

Then $(\alpha', \beta', N) \in \mathcal{A} = A_L^*$ where $N < n'$.

Note: It is possible to show that $\mathcal{A}$ is complete using a different approach. If we define $S = \cup_{i=0}^* S_i$ and S^* as the convex hull of S, it can be shown that S^* is closed. This implies that $\mathcal{A}$ is complete. However, to prove that S^* is closed requires a technique similar to that used in proving Theorem 1.

THEOREM 2. *If $f(x, \theta_0) = 0$ if and only if $f(x, \theta_1) = 0$, then a necessary and sufficient condition for (α, β, n) to belong to $\mathcal{A}$ is that for some non-negative a and b and positive c, we have*

$$a\alpha + b\beta + cn = \min_{(\alpha',\beta',n') \in A^*} \{a\alpha' + b\beta' + cn'\}.$$

PROOF. To prove the sufficiency of the condition, we consider 4 cases.

Case 1. $a = 0, b = 0, c > 0$. Then $a\alpha' + b\beta' + cn' = cn'$ is minimized *only* by tests $(\alpha, \beta, 0)$ belonging to A_0. However, $A_0 \subset \mathcal{A}$, proving the sufficiency of the condition if Case 1 holds.

Case 2. $a = 0, b > 0, c > 0$. Then, $a\alpha' + b\beta' + cn' = b\beta' + cn'$ is minimized only by the test $(1, 0, 0)$ which belongs to A_0.

Case 3. $a > 0, b = 0, c > 0$. (Similar to Case 2.)

Case 4. $a > 0, b > 0, c > 0$. Then, it is well known, and can be easily proved that any test (α, β, n) such that $a\alpha + b\beta + cn = \min_{(\alpha',\beta',n') \in A^*}(a\alpha' + b\beta' + cn')$ belongs to $\mathcal{A}$.

To prove the necessity of the condition, we assume $(\alpha, \beta, n) \in \mathcal{A}$.

(i) If $n = 0$, choose $a = 0, b = 0, c = 1$.

(ii) If $n > 0$, then it is well known in the theory of convex sets that there exist non-negative numbers a, b and c such that

$$a\alpha + b\beta + cn = \min_{(\alpha',\beta',n') \in A^*} (a\alpha' + b\beta' + cn').$$

It remains to show that $c > 0$. Assume $c = 0$. Then

$$a\alpha + b\beta = \min_{(\alpha',\beta',n') \in A^*} (a\alpha' + b\beta') = 0.$$

Since $(\alpha, \beta, n) \varepsilon \mathcal{A}$, then there exist numbers $\gamma_i, \alpha_i, \beta_i, n_i$ such that $(\alpha, \beta, n) = \sum_{i=1}^{4} \gamma_i(\alpha_i, \beta_i, n_i)$, where $\gamma_i \geqq 0, \sum_{i=1}^{4} \gamma_i = 1$, $(\alpha_i, \beta_i, n_i) \varepsilon A$ for $i = 1, 2, 3, 4$.
Thus,

$$a\alpha + b\beta = a \sum_{i=1}^{4} \gamma_i \alpha_i + b \sum_{i=1}^{4} \gamma_i \beta_i = 0.$$

Since both a and b cannot equal 0, either $\alpha = 0$ or $\beta = 0$. Assume $\alpha = 0$. Then, if $\gamma_i > 0$, $\alpha_i = 0$. Using the fact that $f(x, \theta_0) = 0$ if and only if $f(x, \theta_1) = 0$, it follows that if $\alpha_i = 0$, $\beta_i = 1$. Hence, $(\alpha, \beta, n) = (0, 1, n)$. But, $(0, 1, n) \not\varepsilon \mathcal{A}$ since (0, 1, 0) is preferred. Thus we are led to a contradiction of the fact that $(\alpha, \beta, n) \varepsilon \mathcal{A}$. If we assume $\beta = 0$, we are led to a similar contradiction. Therefore, the assumption $c = 0$ is false. Theorem 2 is thus proved.

Theorem 2 states, in effect, that the problem of generating $\mathcal{A}$ reduces to constructing tests (α, β, n) which minimize the expression $a\alpha + b\beta + cn$ for all choices of non-negative a and b and positive c. The cases where either a or b is 0 were discussed and disposed of in proving Theorem 2. The main problem, then, is to construct the tests (α, β, n) which minimize the expression $a\alpha + b\beta + cn$. We proceed as follows: without any loss of generality we may assume that $a + b = 1$ and write $a = \pi$ and $b = 1 - \pi$, where $0 < \pi < 1$. Then, we wish to find the tests (α, β, n) in $\mathcal{A}$ such that

$$\pi\alpha + (1 - \pi)\beta + cn = \min_{(\alpha', \beta', n') \varepsilon A^*} [\pi\alpha' + (1 - \pi)\beta' + cn'].$$

Clearly,

$$\min_{(\alpha', \beta', n') \varepsilon A^*} [\pi\alpha' + (1 - \pi)\beta' + cn']$$

$$= \min_{\substack{[\gamma_i, \alpha_i, \beta_i, n_i : \gamma_i \geqq 0, \sum_{i=0}^{\infty} \gamma_i = 1, (\alpha_i, \beta_i, n_i) \varepsilon A] \\ i = 0, 1, 2, \cdots}} \cdot \left\{\pi \sum_{i=0}^{\infty} \gamma_i \alpha_i + (1 - \pi) \sum_{i=0}^{\infty} \gamma_i \beta_i + c \sum_{i=0}^{\infty} \gamma_i n_i\right\}$$

$$= \min_{\substack{[\gamma_i, \alpha_i, \beta_i, n_i : \gamma_i \geqq 0, \sum_{i=0}^{\infty} \gamma_i = 1, (\alpha_i, \beta_i, n_i) \varepsilon A] \\ i = 0, 1, 2, \cdots}} \cdot \left\{\sum_{i=0}^{\infty} \gamma_i [\pi\alpha_i + (1 - \pi)\beta_i] + c \sum_{i=0}^{\infty} \gamma_i n_i\right\}$$

$$= \min_{N \geqq 0} \left(cN + \left\{ + \min_{(\gamma_i, n_i, \sum_{i=0}^{\infty} \gamma_i n_i = N, \gamma_i \geqq 0, \sum \gamma_i = 1, i = 0, 1, 2, \cdots)} \cdot \sum_{i=0}^{\infty} \gamma_i \min_{(\alpha_i, \beta_i, n_i) \varepsilon A_{n_i}} [\pi\alpha_i + (1 - \pi)\beta_i]\right\}\right)$$

It should be noted that the operation "min"$_{N \geqq 0}$ is not restricted to integral values of N.

From the above, it is clear that the desired minimization can be accomplished in 3 steps, which we shall now describe in detail.

Step 1. We can, for each n_i, find the tests (α_i, β_i, n_i) belonging to A_{n_i} which minimize the expression $\pi\alpha_i + (1 - \pi)\beta_i$. For each n_i, let

$$R_\pi(n_i) = \min_{[(\alpha_i,\beta_i):(\alpha_i,\beta_i,n_i)\varepsilon A_{n_i}]} \{\pi\alpha_i + (1 - \pi)\beta_i\}.$$

$R_\pi(n_i)$ may be interpreted as the Bayes risk for fixed sample-size procedures of sample size n_i where π is the a priori probability that θ_0 is the true parameter and $1 - \pi$ is the a priori probability that θ_1 is the true parameter.

In particular,

$$R_\pi(0) = \min_{[\alpha,\beta:0\leqq\alpha\leqq1,\alpha+\beta=1]} \{\pi\alpha + (1 - \pi)\beta\} = \min(\pi, 1 - \pi).$$

If $0 < \pi < \frac{1}{2}$, $R_\pi(0) = \pi$. The only test $(\alpha, \beta, 0)$ belonging to A_0 satisfying the equation $\pi\alpha + (1 - \pi)\beta = \pi$ is the test $(1, 0, 0)$. Similarly, if $\frac{1}{2} < \pi < 1$, $R_\pi(0) = 1 - \pi$. The only test belonging to A_0 satisfying the equation $\pi\alpha + (1 - \pi)\beta = 1 - \pi$ is the test $(0, 1, 0)$. If $\pi = \frac{1}{2}$, $R_\pi(0) = \frac{1}{2}$. Then, any test belonging to A_0 satisfies the equation $\frac{1}{2}\alpha + \frac{1}{2}\beta = \frac{1}{2}$, since $\alpha + \beta = 1$.

We note that

$$\min_{(\alpha',\beta',n')\varepsilon A^*} [\pi\alpha' + (1 - \pi)\beta' + cn']$$

$$= \min_{N\geqq0}\left\{cN + \min_{[\gamma_i,n_i:\gamma_i\geqq0,\sum_{i=0}^\infty\gamma_i=1,\sum_{i=0}^\infty\gamma_i n_i=N]} \sum_{i=0}^\infty \gamma_i R_\pi(n_i)\right\}.$$

Step 2. Subject to the conditions

$$\gamma_i \geqq 0, i = 0, 1, 2, \cdots, \sum_{i=0}^\infty \gamma_i = 1, \sum_{i=0}^\infty \gamma_i n_i = N,$$

we can, for each non-negative value of N choose the γ_i's so that $\sum_{i=0}^\infty \gamma_i R_\pi(n_i)$ is minimized. To this end, let

$$R_\pi = \bigcup_{k=0}^\infty (k, R_\pi(k)).$$

Let R_π^* denote the convex hull of R_π and let $\mathcal{R}_\pi$ denote the lower boundary of R_π^*, i.e., $\mathcal{R}_\pi = \{(k, r)$: (a) $(k, r)\ \varepsilon\ R_\pi^*$ and (b) there exists no point (k', r') belonging to R_π^* such that $k' \leqq k$, $r' \leqq r$, at least one of these inequalities being strict.$\}$.

Then, to accomplish Step 2 of the minimization, given $N \geqq 0$, we merely select the point (N, r) belonging to $\mathcal{R}_\pi$. Since (N, r) is a boundary point of a two dimensional convex set, *(N, r) can always be expressed as a convex linear combination of at most two points in R.* We define

$$r_\pi(N) = \min_{[\gamma_i,n_i:\gamma_i\geqq0,\sum_{i=0}^\infty\gamma_i=1,\sum_{i=0}^\infty\gamma_i n_i=N, i=0,1,2,\cdots]} \left\{\sum_{i=0}^\infty \gamma_i R_\pi(n_i)\right\}.$$

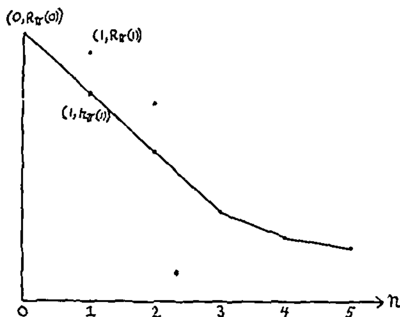


FIG. 2

See Fig. 2. We note that

$$\min_{(\alpha',\beta',n')\in A^*} [\pi\alpha' + (1-\pi)\beta' + cn'] = \min_{N\geq 0} [r_\pi(N) + cN].$$

Step 3. We now wish to choose $N \geq 0$ to minimize the expression $r_\pi(N) + cN$. Since $r_\pi(N)$ is a strictly decreasing, convex and piecewise linear function of N, there exists at least one value of N and at most a finite interval of values of N which minimize $r_\pi(N) + cN$.

It should be noted that if we are given a specific value of N, then there exists a number $c > 0$ such that $r_\pi(N) + cN = \min_k [r_\pi(k) + ck]$. Therefore, for an arbitrary but fixed value of $N > 0$ any procedure obtained in Step 2 will be an admissible mixed single sample test so that Step 3 is inessential in constructing $\mathfrak{a}$.

We shall apply the technique in several problems in the following sections

3. Testing the mean of a normal distribution when the variance is known. Let

$$f(x,\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\},$$

where $\sigma > 0$ is known. We wish to test the hypothesis $H_0:\theta = \theta_0$ against the alternative $H_1:\theta = \theta_1$, $\theta_0 < \theta_1$. It can be shown that for any integer $n \geq 0$,

$$A_n = \{(\alpha,\beta,n):\alpha = 1 - \Phi(t), \beta = \Phi(t - \sqrt{n}\delta) \text{ for } -\infty \leq t \leq \infty\},$$

where

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

and

$$\delta = \frac{\theta_1 - \theta_0}{\sigma}.$$

Step 1. We have already seen that $R_\pi(0) = \min(\pi, 1 - \pi)$. For any integer $n > 0$,

$$R_\pi(n) = \min_{\{(\alpha,\beta):(\alpha,\beta,n)\in A_n\}} [\pi\alpha + (1 - \pi)\beta]$$

$$(2) \qquad = \min_t \{\pi[1 - \Phi(t)] + (1 - \pi)\Phi(t - \sqrt{n}\delta)\}$$

$$= \pi\left[1 - \Phi\left(\frac{\xi}{\sqrt{n}\delta} + \frac{\sqrt{n}\delta}{2}\right)\right] + (1 - \pi)\,\Phi\left(\frac{\xi}{\sqrt{n}\delta} - \frac{\sqrt{n}\delta}{2}\right)\Bigg],$$

where $\xi = \log \pi/(1 - \pi)$. Furthermore, the test (α, β, n) such that

$$\alpha = 1 - \Phi\left(\frac{\xi}{\sqrt{n}\delta} + \frac{\sqrt{n}\delta}{2}\right) \quad \text{and} \quad \beta = \Phi\left(\frac{\xi}{\sqrt{n}\delta} - \frac{\sqrt{n}\delta}{2}\right)$$

is unique. It should also be noted that for any π such that $0 < \pi < 1$, $R_\pi(n)$ is a strictly decreasing function of n. See Figure 3.

Step 2. To accomplish Step 2 of the minimization, we consider $R_\pi(n)$ formally as a function of a continuous variable n. We shall first show that there exists a number $n_i = n_i(\pi)$ such that $R_\pi(n)$ is concave on the interval $(0, n_i)$ and convex on the interval (n_i, ∞). To show the existence of n_i, we use the identities

(a) $\varphi(x - y) = e^{2xy}\varphi(x + y)$,

(b) $\varphi'(x) = -x\varphi(x)$, where $\varphi(x) = \Phi'(x)$.

A routine calculation shows that

(c) $R_\pi'(n) = \dfrac{-\pi}{2\sqrt{n}\delta}\,\varphi\left(\dfrac{\xi}{\sqrt{n}\delta} + \dfrac{\sqrt{n}\delta}{2}\right)$

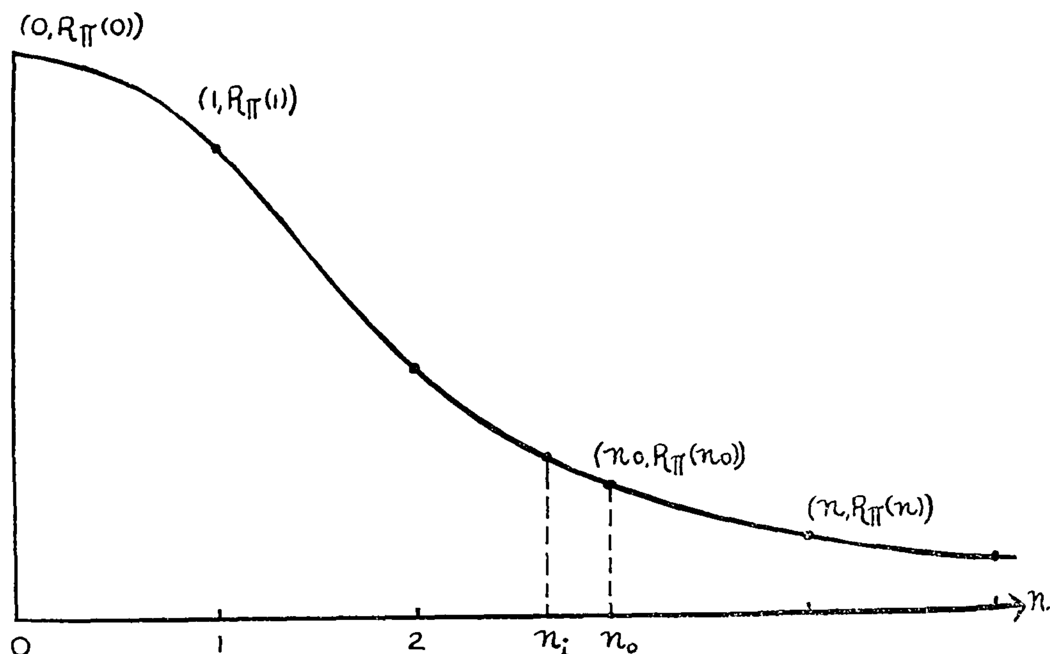


FIG. 3

and

(d) $R''_\pi(n) = (n^2\delta^4 + 4n\delta^2 - 4\xi^2)\dfrac{\pi\sqrt{n}\delta}{16n^3}\varphi\left(\dfrac{\xi}{\sqrt{n}\delta} + \dfrac{\sqrt{n}\delta}{2}\right).$

Setting $R''_\pi(n)$ equal to 0, we find that

$$n_i = \frac{-2 + 2\sqrt{1+\xi^2}}{\delta^2}. \tag{3}$$

Therefore, (3) gives a unique inflection point of the function $R_\pi(n)$. See Fig. 3.

Since $R_\pi(n)$, defined in (2), is defined only for integral values of n, and since n_i in general is not an integer, we assert that there exists an integer $n_0 = n_0(\pi)$ such that $R_\pi(n)$ is concave on the interval $(0, n_0)$ and convex on the interval (n_0, ∞). See Fig. 3. It then follows that

$$r_\pi(N) = \begin{cases} \left(1 - \dfrac{N}{n_0}\right)R_\pi(0) + \dfrac{N}{n_0}R_\pi(n_0) & \text{if } N \leqq n_0 \\ ([N] + 1 - N)R_\pi([N]) + (N - [N])R_\pi([N] + 1) & \text{if } N > n_0 \end{cases}$$

Thus Step 2 of the minimization is achieved.

It now becomes clear that improved randomized procedures (α, β, n) exist and are of the form $(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$ or $(\alpha, \beta, n) = \gamma(1, 0, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$ where $0 < \gamma < 1$ and where

$$n_0 = n_0(\pi), \qquad \alpha_0 = \alpha_0(\pi) = 1 - \Phi\left(\frac{\xi}{\sqrt{n_0}\delta} + \frac{\sqrt{n_0}\delta}{2}\right),$$

$$\beta_0 = \beta_0(\pi) = \Phi\left(\frac{\xi}{\sqrt{n_0}\delta} - \frac{\sqrt{n_0}\delta}{2}\right)$$

for some π such that $0 < \pi < 1$.

It also becomes clear that a test $(\alpha, \beta, n) \varepsilon A_n$ if and only if $n \geqq n_0(\pi)$, where π is defined by the equation

$$\beta = \Phi\left(\frac{\log\dfrac{\pi}{1-\pi}}{\sqrt{n}\delta} - \frac{\sqrt{n}\delta}{2}\right).$$

This gives a complete answer to the general question of whether or not a fixed sample size procedure can be improved upon by means of randomization.

3.1. We now consider the following problem: Given α and β, how can we find the test in $\mathfrak{A}$ achieving the given α and β? To this end, consider two cases.

Case 1. $\alpha < \beta$. Let $\mathfrak{A}_0 = \{(\alpha_0, \beta_0, n_0): n_0 = n_0(\pi), \beta_0 = \beta_0(\pi), \alpha_0 = \alpha_0(\pi)$ for $\frac{1}{2} < \pi < 1\}$.

Let (α, β, n) denote the test in $\mathfrak{A}$ with the given α and n.

From the discussion of Step 2, it is evident that (α, β, n) is an improved randomized procedure if and only if $(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$, where $0 < \gamma < 1$ and where $(\alpha_0, \beta_0, n_0) \varepsilon \mathfrak{A}$. In this case, $\alpha = (1 - \gamma)\alpha_0$, $\beta =$

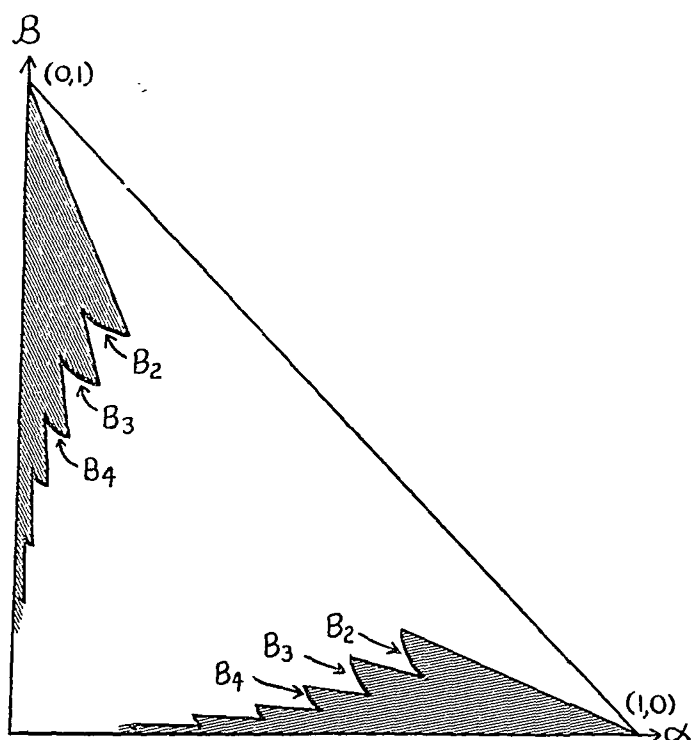


FIG. 4. Shaded region corresponds to the set of (α, β) for which the admissible test (α, β, n) is a randomized procedure; $B_i = \{(\alpha_0, \beta_0): n_0 = i\}$, $\bigcup_{i=2}^{\infty} B_i = P(\mathcal{A}_0 \mid \alpha, \beta)$.

$\gamma + (1 - \gamma)\beta_0$, $n = (1 - \gamma)n_0$. These equations imply that $\alpha/(1 - \beta) = \alpha_0/(1 - \beta_0)$ and $1 - \gamma = \alpha/\alpha_0$. The equation $\alpha/(1 - \beta) = \alpha_0/(1 - \beta_0)$ when interpreted geometrically means that the points $(0, 1)$, (α, β) and (α_0, β_0) are colinear. The equation $1 - \gamma = \alpha/\alpha_0$ when interpreted geometrically means that (α, β) is between $(0, 1)$ and (α_0, β_0).

If (α, β, n) is not an improved randomized procedure, then

$$(\alpha, \beta, n) = \gamma(\alpha_1, \beta_1, [n]) + (1 - \gamma)(\alpha_2, \beta_2, [n] + 1)$$

where $0 \leqq \gamma \leqq 1$ and where $(\alpha_1, \beta_1, [n])$ and $(\alpha_2, \beta_2, [n] + 1) \varepsilon A$ and is of little interest.

We summarize the preceding as follows: Let $P(\mathcal{A}_0 \mid \alpha, \beta)$ denote the projection of $\mathcal{A}_0$ on the (α, β) plane. See Fig. 4. It was convenient to let $\delta = 1$. If (α, β) lies on a line segment joining $(0, 1)$ to one of the points (α_0, β_0) in $P(\mathcal{A}_0 \mid \alpha, \beta)$, then the test $(\alpha, \beta, n) = (1 - \alpha/\alpha_0)(0, 1, 0) + \alpha/\alpha_0(\alpha_0, \beta_0, n_0)$ is the test in $\mathcal{A}$ with the given α and β. Otherwise, (α, β, n) is achieved by randomizing over two fixed sample size procedures, one in $A_{[n]}$ and the other in $A_{[n]+1}$.

Case 2. $\alpha > \beta$. Similar to Case 1.

Table (1) shows the improvement in the expected sample size N which can be achieved for selected tests (α, β, n) belonging to $A_n - \mathcal{A}$. In this case, we let $\delta = .1$.

3.2. Consider next the following problem: Given α and n, how can we construct the test in $\mathcal{A}$ having the given α and n? We solve this problem geometri-

TABLE 1

α	β	Sample size, n, of admissible single sample tests achieving the given α and β	Expected sample size, N, of admissible mixed single sample test achieving the given α and β	Percent saving $\frac{n - N}{n} \times 100$
005	.862	221	119	46
005	.732	383	287	25
01	.732	147	84	43
01	.463	585	574	2
05	687	134	123	8
05	868	28	20	28

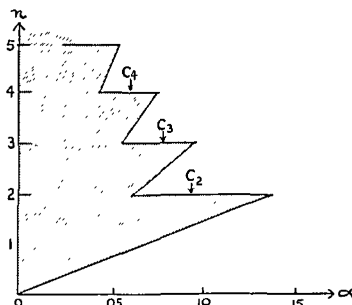


FIG 5 Shaded region corresponds to the set of (α, n) for which the admissible test [illegible] is a randomized procedure; $C_i = \{(\alpha_0, n_0): n_0 = i\}$, $\bigcup_{i=2}^{\infty} C_i = P(\mathfrak{A}_0 \mid \alpha, n)$

cally. Let $P(\mathfrak{A}_0 \mid \alpha, n)$ denote the projection of the set $\mathfrak{A}_0$ on the $(\alpha,$ [illegible] See Fig. 5. Then, draw a line of slope n/α through the origin. Determ[illegible] of intersection (α_0, n_0) of this line and $P(\mathfrak{A}_0 \mid \alpha, n)$. Clearly, [illegible] I[illegible] $\alpha_0 > \alpha$, the test in $\mathfrak{A}$ having the given α and n is the mixture

$$\frac{\alpha}{\alpha_0}(\alpha_0, \beta_0, n_0) + \left(1 - \frac{\alpha}{\alpha_0}\right)(0, 1, 0).$$

If $\alpha_0 \leqq \alpha$, the test in $\mathfrak{A}$ having the given α and n is a mixt[illegible] $A_{[n]}$ and the other in $A_{[n]+1}$ and hence is of little inte[illegible]

4. Tests on the mean of a binomial distribution. L[illegible]

$$f(x, \theta) = \theta^x(1 - \theta)^{1}$$ [illegible]

$$= 0$$ [illegible]

We wish to test the hypothesis $H_0:\theta = \theta_0$ against the alternative $H_1:\theta = \theta_1$, $\theta_1 > \theta_0$. It is known that

$$A_n = \Big\{(\alpha, \beta, n): \alpha = \sum_{i=0}^{n+1} \gamma_i \alpha_i , \beta = \sum_{i=0}^{n+1} \gamma_i \beta_i ,$$

$$\gamma_0 = \gamma_1 = \gamma_{i-1} = \gamma_{i+2} = \cdots \gamma_{n+1} = 0, \gamma_i \geqq 0,$$

$$\gamma_{i+1} \geqq 0, \sum_{i=0}^{n+1} \gamma_i = 1,$$

$$\alpha_i = \sum_{i=1}^{n} \binom{n}{r} \theta_0^r (1 - \theta_0)^{n-r}, \qquad \beta_i = \sum_{r=0}^{i=1} \binom{n}{r} \theta_1^r (1 - \theta_1)^{n-r},$$

$$i = 0, 1, 2, \cdots n + 1\Big\}.$$

Howard Raiffa [2] has pointed out that if we consider the projections of A_1 and A_2 on the (α, β) plane, there exists a test in A_2 whose operating characteristic is $(\theta_0 , 1 - \theta_1 , 2)$. However, there exists a test in A_1 whose operating characteristic is $(\theta_0 , 1 - \theta_1 , 1)$. Hence $(\theta_0 , 1 - \theta_1 , 2) \not\in \mathfrak{A}$. See Fig. 6. Furthermore, if π is such that

$$\frac{\pi}{1 - \pi} = \frac{\theta_0(1 - \theta_0)}{\theta_1(1 - \theta_1)},$$

then $R_\pi(1) = R_\pi(2)$.

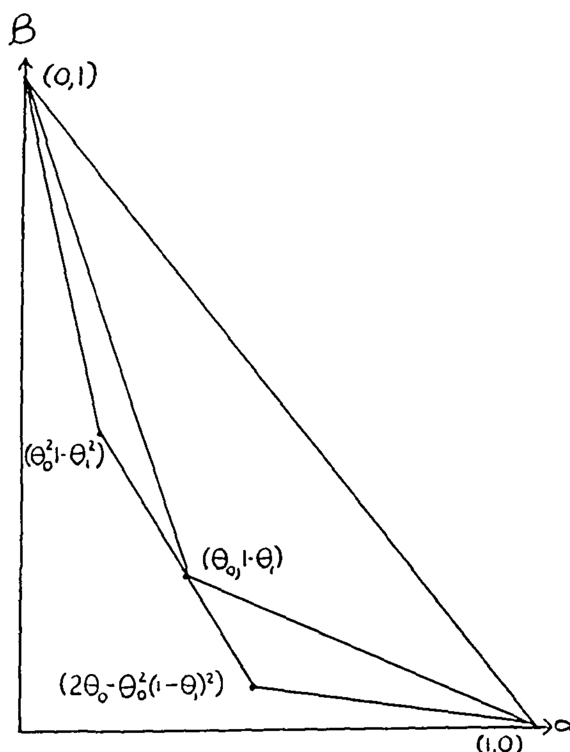


FIG. 6

TABLE 2

α	n	Probability of a type II error of admissible single sample test having the given α and n	Probability of a type II error of randomized test having the given α and n	Percent decrease
.512	30	.029	.018	38
.361	20	.112	.090	19
.098	20	.320	.286	11
.350	40	.030	.024	20

Unlike the normal distribution, there does not exist an integer $n_0(\pi)$ such that $R_\pi(n)$ is concave on the interval $(0, n_0(\pi))$ and convex on the interval $(n_0(\pi), \infty)$. Rather, it was found by numerical calculation that $R_\pi(n)$ has many inflection points. Thus, we do not generalize any further and present the following examples.

Example I. Let $\theta_0 = .04$, $\theta_1 = .15$. Table 2 shows the percent decrease in the probability of a type II error that randomization achieves over fixed sample size procedures for the given α and n. Since $R_\pi(n)$ was calculated for values of n where $n = 5k$ where k is a non-negative integer, it cannot be said with certainty that the improvements shown in Table 2 are optimal. However, the optimal improvements are at least as great as the ones recorded.

Example II. We again wish to test the hypothesis $H_0{:}\theta = \theta_0$ against the alternative $H_1{:}\theta = \theta_1$ where $\theta_0 < \theta_1$. Then, it is well known that any test $(\alpha, \beta, 1)$ such that

$$(\alpha, \beta, 1) = \gamma(0, 1, 1) + (1 - \gamma)(\theta_0, 1 - \theta_1, 1),$$

where $0 \leqq \gamma \leqq 1$ belongs to A_1. We shall now show that if we are given a test $(\alpha, \beta, 1)$ of the above type such that $\gamma \leqq 1 - \theta_1/2$ then there exists a mixed single sample test $(\alpha^*, \beta, 1)$ such that

$$\frac{\alpha - \alpha^*}{\alpha} = \frac{\gamma(1 - \alpha - \beta)}{(1 - \gamma)(1 + \beta - 2\gamma)}. \tag{4}$$

The expression $(\alpha - \alpha^*)/\alpha$ is interpreted as the fractional saving [illegible] by randomization.

To prove this, consider the test

$$(\alpha^*, \beta', n') = \frac{\gamma}{2 - \theta_1}(0, 1, 0) + \frac{\gamma}{2 - \theta_1}(\theta_0^2, 1 - \theta_1^2, 2)$$

$$+ \Big(1 - [illegible]$$

Since $\gamma \leqq 1 - \theta_1/2$, the above test is a bonafide mixture [illegible] $\beta' = \beta$, $n' = 1$ and that $(\alpha - \alpha^*)/\alpha$ has the value given [illegible]

To illustrate the fractional saving in α which can [illegible]

test $H_0:\theta = .10$ against $H_1:\theta = .95$. Then, there exists a test $(\alpha, \beta, 1)$ in A_1 where $(\alpha, \beta, 1) = .5(0, 1, 1) + .5(.10, .05, 1) = (.05, .525, 1)$. Consider the test

$$(\alpha^*, \beta, 1) = \frac{.5}{2 - .95}(0, 1, 0) + \frac{.5}{2 - .95}(.01, .0975, 2) + \left(1 - \frac{1}{2 - .95}\right)(.10, .05, 1) = \left(\frac{1}{105}, .525, 1\right).$$

Then

$$\frac{\alpha - \alpha^*}{\alpha} = \frac{17}{21}.$$

5. Tests on the range of a rectangular distribution when one endpoint is known. Let

$$f(x, \theta) = \frac{1}{\theta} \qquad \text{if } 0 \leqq x \leqq \theta,$$
$$= 0 \qquad \text{elsewhere.}$$

We wish to test the hypothesis $H_0:\theta = \theta_0$ against the alternative $H_1:\theta = \theta_1$, $\theta_1 > \theta_0$. It can be shown that

$$(5)\qquad \begin{aligned} A_n &= \left\{(\alpha, \beta, n):\alpha = \frac{\theta_0^n - t^n}{\theta_1^n}, \beta = \frac{t^n}{\theta_1^n}, 0 \leqq t \leqq \theta_0\right\} \\ &= \left\{(\alpha, \beta, n):0 \leqq \alpha \leqq 1, \beta = \left(\frac{\theta_0}{\theta_1}\right)^n (1 - \alpha)\right\}. \end{aligned}$$

It should be noted that Theorem 2 does not hold since $f(x, \theta_0)$ and $f(x, \theta_1)$ do not vanish simultaneously for values of x such that $\theta_0 \leqq x \leqq \theta_1$. Hence, we shall alter our approach to generating $\mathfrak{A}$ by proving a theorem which will yield as a consequence a technique for constructing $\mathfrak{A}$.

THEOREM 3. *If $(\alpha, \beta, n) \in A_n$ where $\alpha > 0$ and $n > 0$, then $(\alpha, \beta, n) \notin \mathfrak{A}$.*

PROOF. If $(\alpha, \beta, n) \in A_n$, then it follows from (5) that $\beta = (\theta_0/\theta_1)^n(1 - \alpha)$. Consider the test

$$\begin{aligned} (\alpha', \beta', n') &= \alpha(1, 0, 0) + (1 - \alpha)\left(0, \left(\frac{\theta_0}{\theta_1}\right)^n, n\right) \\ &= \left(\alpha, (1 - \alpha)\left(\frac{\theta_0}{\theta_1}\right)^n, (1 - \alpha)n\right) \\ &= (\alpha, \beta, (1 - \alpha)n). \end{aligned}$$

Clearly $(\alpha, \beta, (1 - \alpha)n)$ is preferred to (α, β, n). Theorem 3 states that all single sample tests (α, β, n) such that $0 < \alpha \leqq 1$ and $n > 0$ are inadmissible in the class of mixed single-sample tests. Consequently, the class $\mathfrak{T}$ can be generated by the test $(1, 0, 0)$ and the sequence of tests $\{(0, (\theta_0/\theta_1)^k, k)\}$, $k = 0, 1, 2, \cdots$. Since $(\theta_0/\theta_1)^n$ is a convex function of n, it can be shown that $(\alpha, \beta, n) \in \mathfrak{A}$ if and only if $(\alpha, \beta, n) = \gamma_1(1, 0, 0) + \gamma_2(0, \theta_0/\theta_1)^k, k) + \gamma_3(0, (\theta_0/\theta_1)^{k+1}, k + 1$ for some non-negative numbers $\gamma_1, \gamma_2, \gamma_3$ and some non-negative integer k where

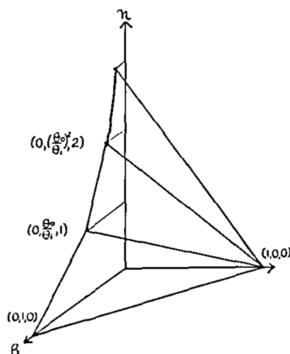


Fig. 7

$\sum_{i=1}^{3} \gamma_i = 1$. In fact it is easily verified that $k = [n/1 - \alpha]$, $\gamma_1 = \alpha$, $\gamma_2 = (1 - \alpha)([n/1 - \alpha] + 1) - n$ and $\gamma_3 = n - (1 - \alpha)[n/1 - \alpha]$. See Fig. 7.

Corollary 1. *If $(\alpha, \beta, n) \varepsilon A_n$, there exists a test $(\alpha, \beta, n') \varepsilon \mathfrak{A}$ where $n' = (1 - \alpha)n$.*

Proof. From the preceding discussion, the test $(\alpha, \beta, n') = \alpha(1, 0, 0) + (1 - \alpha)(0, (\theta_0/\theta_1)^n, n) \varepsilon \mathfrak{A}$. Since $n' = (1 - \alpha)n$, the desired conclusion follows.

We note that the fractional saving in the expected number of observations obtained by randomization is equal to α, *i.e.*,

$$\frac{n - n'}{n} = \frac{n - (1 - \alpha)n}{n} = \alpha.$$

6. Tests on the mean of a rectangular distribution when the range is known. Let

$$f(x, \theta) = 1 \qquad \text{if } \theta < x < \theta + 1,$$
$$= 0 \qquad \text{elsewhere.}$$

We wish to test the hypothesis $H_0: \theta = 0$ against the alternative $H_1: \theta = \theta_1$ where $0 < \theta_1 < 1$. A simple calculation shows that

$$A_n = \{(\alpha, \beta, n): \alpha = (1 - t)^n, \quad \beta = (1 - \theta_1)^n - (1 - t)^n,$$
$$\theta_1 \leqq t \leqq 1\} = \{(\alpha, \beta, n): 0 \leqq \alpha \leqq (1 - \theta_1)^n, \alpha + \beta = (1 - \theta_1)^n\}.$$

See Fig. 8. Let $R_\pi(n) = \min_{(a,b,n)\varepsilon A_n} [\pi\alpha + (1 - \pi)\beta] = \min [\pi(1 - \theta_1)^n, (1 - \pi)(1 - \theta_1)^n] = (1 - \theta_1)^n \min (\pi, 1 - \pi)$. Obviously $R_\pi(n)$ is a convex function of n. It follows that $A_n \subset \mathfrak{A}$. In other words, all fixed sample size tests are admissible in the class of mixed single sample tests.

FIG. 8

7. Confidence interval estimation. We next wish to extend the notion of mixed single sample procedures to confidence interval estimation. Perhaps this purpose can best be served by an illustrative example.

Example. Let X denote a normally distributed random variable with unknown mean μ and known variance σ^2. (There is no loss in generality if we assume that $\sigma^2 = 1$, and we shall do so for the remainder of this section.) We wish to consider the problem of obtaining a confidence interval for μ. The standard procedure consists of

(a) choosing a number α between 0 and 1, called the confidence coefficient.
(b) calculating a number t using the equation $\alpha = 1 - 2\Phi(-t)$.
(c) drawing a sample of n independent observations on X and calculating $\bar{X}$, the sample mean.
(d) making the statement tha the interval $(\bar{X} - t/\sqrt{n}, \bar{X} + t/\sqrt{n})$ covers μ with confidence α.

A confidence interval procedure is evaluated in terms of a triple $(1 - \alpha, L, n)$ where $1 - \alpha$ denotes the probability that the confidence interval will not cover μ, L denotes the length of the confidence interval and n denotes the sample size.

We will now exploit the notion of randomizing over the sample size in confidence interval estimation using an approach similar to the one used in Section 2. For integral values of $n \geqq 1$, we let

$$A_n = \left\{(1 - \alpha, L, n) : \alpha = 1 - 2\Phi(t), L = \frac{2t}{\sqrt{n}}, 0 \leqq t < \infty\right\}$$
$$= \left\{(1 - \alpha, L, n) : \alpha = 1 - 2\Phi\left(-\frac{\sqrt{n}L}{2}\right)\right\}.$$

We define $A_0 = \{(1, 0, 0)\}$.

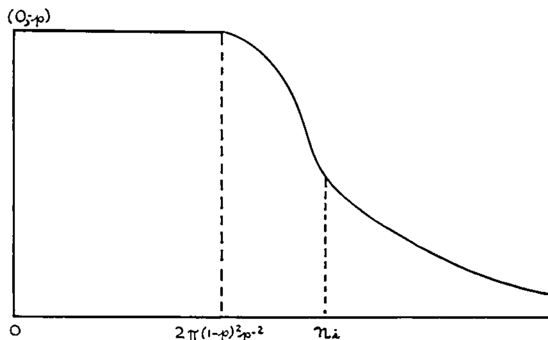


FIG. 9

As an analogue of the Bayes risk $R_\pi(n)$, we consider

$$R_p(n) = \min_{(1-\alpha, L, n) \in A_n} [p(1-\alpha) + (1-p)L].$$

A routine calculation shows that

$$R_p(n) = p \qquad \text{if } 0 \leqq n \leqq 2\pi(1-p)^2 p^{-2},$$

$$= 2p\Phi\left(-\sqrt{\log \frac{np^2}{2\pi(1-p)^2}}\right) + \frac{2(1-p)}{n}\sqrt{\log \frac{np^2}{2\pi(1-p)^2}}$$

$$\text{if } n > 2\pi(1-p)^2 p^{-2},$$

See Fig. 9.

If we treat $R_p(n)$ as a function of a continuous variable n, we find that

$$R_p'(n) = 0 \qquad \text{if } n < \frac{1}{c},$$

$$R_p''(n) = -\frac{(1-p)}{2}\frac{1-3(\log cn)}{\sqrt{n^5}\sqrt{\log cn}} \qquad \text{if } n > \frac{1}{c},$$

where $c = p^2/[2\pi(1-p)^2]$. As in Section 3, there exists a non-negative number $n_i = n_i(p)$ such that $R_p(n)$ is concave on the interval $(0, n_i)$ and convex on the interval (n_i, ∞). In fact, $n_i = [2\pi(1-p)^2]/p^2e^{\frac{1}{3}}$. Using an argument similar to the one used in Section 3, it becomes clear that "improved" mixed confidence interval procedures exist and are of the form

$$(\alpha, L, n) = \gamma(0, 0, 0) + (1-\gamma)(\alpha', L', n'),$$

where $0 < \gamma$ 1 and (α', L', n') is a fixed sample size confidence interval procedure.

TABLE 3

Confidence coefficient α	Expected sample size n	Length of fixed sample size procedure having the given α and n	Expected length of randomized confidence interval procedure having the given α and n	Percent decrease in the expected length
.044	1	.110	.037	66
.392	9	.334	.329	1
.174	4	.220	.146	34

Table 3 gives some examples of admissible mixed single sample procedures and improvements which can be obtained in the expected length of a confidence interval if a mixing scheme is used.

Improved randomized confidence intervals are of such a nature that certain questions are brought to mind. First, how much "confidence" can we place in randomized confidence intervals? It is true that a confidence interval of the form $(\alpha, L, n) = \gamma(0, 0, 0) + (1 - \gamma)(\alpha', L', n')$ will cover μ 100 α% of the time, will have average length L and will have expected sample size n. However, if we are given confidence interval $(0, 0, 0)$, we no longer have confidence α that we are covering μ. On the other hand, if we are given the confidence interval $(\bar{X} - L'/2, \bar{X} + L'/2)$, we have confidence $\alpha' > \alpha$ that we are covering μ. Furthermore, if a statistician uses a mixed procedure and does not tell this to his customers, then his customers can have confidence α—unless, of course, they are given the procedure $(0, 0, 0)$. (However, if we restrict ourselves to procedures where the sample size n is at least 1, then they could still have confidence α.) In other words, by withholding information from his customers, the statistician gives them confidence α. By giving them information, he either reduces their confidence to 0, or increases their confidence to α'.

This is not the only example of such a situation in statistical techniques. Take, for example, the Stein two sample procedure for finding a confidence interval (of fixed length 1 and confidence coefficient α) for the mean of a normal distribution with unknown variance. A sample of n_0 observations is taken and the sample variance S_0^2 is calculated. Then, an additional n_1 observations are taken where

$$n_1 = \max\left\{n_0, \left[\frac{S_0^2}{d}\right] + 1\right\} - n_0,$$

where d depends on α and 1. The two samples are then combined, the mean $\bar{X}$ of the combined samples is calculated and the confidence interval $\left(\bar{X} - \frac{l}{2}, \bar{X} + \frac{l}{2}\right)$ is given. Now, if it turns out that the variance S^2 of the combined samples is much larger than S_0^2, one is led to believe that the second sample size was not large enough. Thus, one's confidence of α might be reduced, given this information. However, if one did not have this information about S^2, then one's confidence would still be α. This situation is indeed similar to the preceding one.

Another peculiarity of mixed single sample confidence interval procedures is that we get short length only when we do not cover μ. This immediately brings to

mind the question of average length as a criterion for a confidence interval procedure. It is clear that small length is desirable if μ is being covered. What one wants when μ is not covered is open to question. Clearly, we can agree that procedures which give small length when μ is not covered and large length when μ is covered are not desirable ones. Randomized procedures are of this nature.

8. The k decision problem. Let X denote a random variable with distribution function $F(x, \theta)$. Instead of considering only two possible values of θ, θ_0 and θ_1, as we did in the previous section, we now consider k possible values of θ. Let $\theta_1, \theta_2, \cdots, \theta_k$ denote the k possible values of θ. We assume that $\theta_1 < \theta_2 < \cdots < \theta_k$. For any fixed sample size decision rule δ_n, based on samples of size n, let $\alpha_i(\delta_n)$ denote the probability that θ_i will not be selected as the true value of θ when θ_i is the true value of θ if the decision rule δ_n is used. Every fixed sample size decision rule is then identified with an operating characteristic $(\alpha_1, \alpha_2, \cdots \alpha_k, n)$ where $\alpha_i = \alpha_i(\delta_n)$ for $i = 1, 2, \cdots, k$ and where n denotes the sample size. The classes S_n, A_n, A, A^* and $\mathfrak{A}$ are defined in an obvious way and the functions $R_\pi(n)$ and $r_\pi(n)$ are defined as in Section 2 where $\pi = (\pi_1, \pi_2, \cdots \pi_k)$, $\pi_i \geqq 0$ and $\sum_{i=1}^k \pi_i = 1$. We can then extend all the results obtained in Section 2 to the k decision problem.

In the particular case

$$F(x, \theta) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{t-\theta}{\sigma}\right)^2\right\} dt,$$

where $\sigma > 0$ is known, we shall show that it is possible to obtain improvements by randomization. For each positive integral value of n, an essentially complete class of decision rules, C_n, can be generated in the following way: Let $(x_1, x_2, \cdots x_n)$ denote a sample of n independent observations on X and let let $(t_0, t_1, \cdots, t_k)$ denote a partition of the real line such that $t_i \leqq t_{i+1}$, $i = 0, 1, \cdots k - 1$. In particular, $t_0 = -\infty$ and $t_k = \infty$. Then any procedure which selects θ_i as the true value of θ whenever $t_{i-1} \leqq \bar{X} < t_i$ is called a monotone procedure. Let C_n denote the class of all monotone procedures The class C_n is known to be essentially complete.

By definition,

$$\begin{aligned} R_\pi(n) &= \min_{(\alpha_1,\alpha_2,\cdots,\alpha_k)\epsilon A_n} \sum_{i=1}^{k} \pi_i\alpha_i = \min_{(\alpha_1,\alpha_2,\cdots,\alpha_k)\epsilon C_n} \sum_{i=1}^{k} \pi_i\alpha_i \\ &= \min_{(t_1,t_2,\cdots,t_{k-1})} \sum_{i=1}^{k} \pi_i\left[1 - \Phi\left(\sqrt{n}\,\frac{(t_i-\theta_i)}{\sigma}\right) + \Phi\left(\sqrt{n}\,\frac{(t_{i-1}-\theta_i)}{\sigma}\right)\right] \\ &= \sum_{i=1}^{k} \pi_i\left[1 - \Phi\left(\frac{\xi_i}{\sqrt{n}\delta_i} - \frac{\sqrt{n}\delta_i}{2}\right) + \Phi\left(\frac{\xi_{i-1}}{\sqrt{n}\delta_{i-1}} + \frac{\sqrt{n}\delta_{i-1}}{2}\right)\right], \end{aligned}$$

where

$$\xi_i = \log\frac{\pi_{i+1}}{\pi_i}, \qquad \delta_i = \frac{\theta_i - \theta_{i+1}}{\sigma}, \qquad i = 1, 2, \cdots, k-1,$$

$$\frac{\xi_0}{\sqrt{n}\delta_0} + \frac{\sqrt{n}\delta_0}{2} = -\infty \quad \text{and} \quad \frac{\xi_k}{\sqrt{n}\delta_k} - \frac{\sqrt{n}\delta_k}{2} = \infty.$$

Considering $R_\pi(n)$ as a function of a continuous variable n, we find

$$\text{(a)}\quad R'_\pi(n) = \sum_{i=2}^{k} \pi_i \,\rho\left(\frac{\xi_{i-1}}{\sqrt{n}\delta_{i-1}} + \frac{\sqrt{n}\delta_{i-1}}{2}\right) \cdot \frac{\delta_{i-1}}{\sqrt{n}}.$$

$$\text{(b)}\quad R''_\pi(n) = \sum_{i=2}^{k} \frac{-1}{\delta_n^{5/2}\,\delta_{i-1}} \left(\frac{\xi_{i-1}}{\sqrt{n}\delta_{i-1}} + \frac{\sqrt{n}\delta_{i-1}}{2}\right) (\delta_{i-1}^4 n^2 + 4\delta_{i-1}^2 n - 4\xi_{i-1}^2).$$

For each value of $i = 2, 3, \cdots k$, the function $f_i(n) = \delta_{i-1}^4 n^2 + 4\delta_{i-1}^2 n - 4\xi_{i-1}^2$ is a quadratic function of n. Since the only non-negative root of the equation $f_i(n) = 0$ is

$$n_i = \frac{-2 + 2\sqrt{1 + \xi_{i-1}^2}}{\delta_{i-1}^2},$$

it follows that

$$f_i(n) \leqq 0 \qquad \text{if } 0 \leqq n \leqq n_i,$$

$$f_i(n) > 0 \qquad \text{if } n > n_i.$$

Then, since $\delta_{i-1} < 0$ for $i = 2, 3, \cdots, k$, it follows that

$$R''_\pi(n) < 0 \qquad \text{if } n < \min_i (n_i)$$

and

$$R''_\pi(n) > 0 \qquad \text{if } n > \max_i (n_i)$$

Hence, if we let $a = \min_i (n_i)$ and $b = \max_i (n_i)$, it follows that $R_\pi(n)$ is concave on the interval $(0, a)$ and convex on the interval (b, ∞). Clearly, $a \leqq b$.

Thus, for certain values of $\pi_1, \pi_2, \cdots \pi_k$ and $\theta_1, \theta_2, \cdots, \theta_k$, it is possible to achieve improvements by randomization.

9. Testing a composite hypothesis against a composite alternative. We next wish to extend the notion of mixed single sample tests to the problem of testing a composite hypothesis againxt a composite alternative. To fix ideas, let

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\,\delta} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\},$$

where $\sigma > 0$ is known. We wish to test the hypothesis $H_0:\theta \leqq \theta_0$ against the alternative $H_1:\theta > \theta_1$, $\theta_1 > \theta_0$. If we are given α and n, the "best" fixed sample size test of level α and size n is obtained by using the best fixed sample size test of $H'_0:\theta = \theta_0$ against $H'_1:\theta = \theta_1$ corresponding to the given α and n. The resultant fixed sample size test has the desirable property that its power function $P(\theta \mid \alpha, n)$ tends to 1 as θ tends to infinity.

Can we construct, for given α and n, a "good" mixed single sample test of level α and *expected* sample size n in an analogous way? Clearly, if the best mixed single sample test of H'_0 against H'_1 is a bona fide mixture, it is not even true that its power function, $P(\theta)$, approaches 1 as θ approaches infinity. For, in this case,

the fixed sample size test (0, 1, 0) will be chosen with probability λ, say, where $0 < \lambda < 1$, so that $P(\theta) \leqq 1 - \lambda$ for all θ.

However, it should be noted that the fact that $P(\theta)$ does not tend to 1 as θ tends to infinity is not always undesirable for we know, in certain cases, that the set of possible values of θ is bounded, e.g., in testing the mean height θ of American soldiers, we know that $\theta \leqq 6$ feet 2 inches. Consequently, a test procedure which does not have high power at $\theta = 7$ feet is not necessarily undesirable.

Finally, we note that if we restrict ourselves to randomizing over fixed sample size tests of sample size $n > 1$, then $P(\theta) \to 1$ as $\theta \to \infty$.

10. Comparison with the Wald Sequential Probability Ratio Test. In general, it is difficult to compare the improvements attainable by using the Wald Sequential Probability Ratio Test with improvements attainable by randomizing over fixed sample size procedures. For, every test will now be identified with a quadruple $(\alpha, \beta, E_{\theta_0}(n), E_{\theta_1}(n))$. $E_{\theta_0}(n)$ and $E_{\theta_1}(n)$ are usually difficult to calculate. However, in the case of mixed single sample tests, $E_{\theta_0}(n) = E_{\theta_1}(n)$ and do not depend on the unknown value of θ. In some special cases it is easy to make a comparison and this we shall do.

Example.

$$f(x, \theta) = \frac{1}{\theta} \qquad \text{if } \theta \leqq x < \theta,$$

$$= 0 \qquad \text{elsewhere.}$$

It can be shown that if we use Wald's test, only two types of tests are attainable. They are the test (1, 0, 0, 0) or tests of the form

$$\left(0, \left(\frac{\theta_0}{\theta_1}\right)^k, k, \frac{1 - \left(\frac{\theta_0}{\theta_1}\right)^k}{1 - \left(\frac{\theta_0}{\theta_1}\right)}\right),$$

where k is a non-negative integer. However, using mixed single sample tests, we can attain the test (1, 0, 0, 0) and tests of the form $(0, (\theta_0/\theta_1)^k, k, k)$ where k is a non-negative integer, and mixtures of such tests. Since

$$\lim_{\frac{\theta_0}{\theta_1} \to 1} \frac{k}{\left[\dfrac{1 - \left(\frac{\theta_0}{\theta_1}\right)^k}{1 - \left(\frac{\theta_0}{\theta_1}\right)}\right]} = 1,$$

it is clear that if θ_0/θ_1 is close to 1, then mixed single sample procedures are almost as good as Wald procedures.

11. Estimation. Can mixing fixed sample estimation procedures yield improvements in estimation techniques? If we evaluate a fixed sample size estimator t_n in terms of a pair of numbers $\{E[L(t_n, \theta)], n\}$, where $E[L(t_n, \theta)]$ denotes the

expected loss if the estimator t_n is used when θ is the true parameter and where n denotes the sample size, then mixing over fixed sample size procedures will not yield improvements since in all problems of practical interest $E[L(t_n, \theta)]$ is a convex function of n. For example, if we wish to estimate the mean θ of a distribution with finite variance σ^2, then, if $t_n = \bar{X}$ and if $L(t_n, \theta) = k(\bar{X} - \theta)^2$, we find that $E[L(t_n, \theta)] = k\, \sigma^2/n$. Thus, it will not pay to randomize.

12. Conclusion. In what situations is a mixed single sample procedure justifiable? In order to answer this question, we must first realize that throughout this paper, we have been judging a test δ by its operating characteristic (α, β, n). If this triple is our only means of evaluating a test procedure, then it is true that single sample procedures would not be justifiable since a sequential probability ratio test achieving the given α and β would be better. However, practical considerations might limit one to a single stage of sampling, e.g., in agricultural experiments, one might not wish to use more than one stage of sampling; or, if one is testing electric light bulbs, one might not wish to test the bulbs sequentially. Other examples could be given.

One could reasonably ask why fixed sample size procedures should not always be used in these situations. Presumably, if the experiment were a so called "one shot affair", i.e., if the experiment were never to be repeated, then one might reasonably insist on a non-randomized fixed sample size procedure (although, of course, this position is not universally held). However, if one repeats the experiment often, it would be reasonable to use a mixed sample size procedure. To illustrate this point, consider Example II in Section 4. In this example, suppose θ_0 represents the probability that a person who has been contaminated with a certain disease will respond positively to a certain test and θ_1 represents the probability that a person who has not been contaminated will respond positively to this same test. Then, if several thousand people are to be classified as either contaminated or non-contaminated according to this test, then the mixed test (1/101, .525, 1) would be preferred to the test (.05, .525, 1) since the mixed test will falsely classify less than 1 percent of the contaminated people whereas the fixed sample size procedure will misclassify 5 percent of the contaminated people. On the other hand, both tests will misclassify the same percentage of non-contaminated people, and both procedures will use on the average of one test per person.

At this point, one could raise strenuous objections to mixed single sample tests on grounds similar to those raised in Section 7, i.e., if one is told which single sample test is actually used, the conditional probabilities of misclassification are no longer α and β. For example, consider a mixed test of the form

$$(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha', \beta', n').$$

Now, suppose that a person is told that he has been classified according to the test (0, 1, 0). Such a person would of course be most unhappy. On the other hand, if he is not told which of the tests was used, he would maintain his con-

fidence in the procedure used. In other words, *by withholding information, one can influence a person's willingness to accept a result.* Some feel that axiomatically this is an untenable policy.

13. Acknowledgements. The author is indebted to Allan Birnbaum for suggesting the Bayes approach exploited in this paper. The author is also indebted to Howard Raiffa for his many helpful suggestions and constructive criticisms.

REFERENCES

[1] W. Kruskal, Private communication.

[2] H Raiffa, "Statistical decision theory approach to item selection for dichotomous test criterion variables," School of Aviation Medicine Series in Statistics, Report No. 56-139, Randolph Field, 1957

[3] A. Birnbaum, "Applications of vector-valued risk functionals," *Ann. Math Stat*, Vol 27 (1956), pp. 544-545 (abstract).

[4] J. L. Hodges, Jr, and E. L. Lehmann, "The efficiency of some non-parametric competitors of the t-test", *Ann Math. Stat*, Vol. 27, No. 2 (1956) pp. 324-335

ASYMPTOTIC NORMALITY AND EFFICIENCY OF CERTAIN NONPARAMETRIC TEST STATISTICS[1]

By Herman Chernoff and I. Richard Savage

Stanford University and *University of Minnesota*

1. Summary. Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be ordered observations from the absolutely continuous cumulative distribution functions $F(x)$ and $G(x)$ respectively. If $z_{Ni} = 1$ when the ith smallest of $N = m + n$ observations is an X and $z_{Ni} = 0$ otherwise, then many nonparametric test statistics are of the form

$$mT_N = \sum_{i=1}^{N} E_{Ni} z_{Ni} .$$

Theorems of Wald and Wolfowitz, Noether, Hoeffding, Lehmann, Madow, and Dwass have given sufficient conditions for the asymptotic normality of T_N. In this paper we extend some of these results to cover more situations with $F \neq G$. In particular it is shown for all alternative hypotheses that the Fisher-Yates-Terry-Hoeffding c_1-statistic is asymptotically normal and the test for translation based on it is at least as efficient as the t-test.

2. Introduction. Finding the distributions of nonparametric test statistics and establishing optimum properties of these tests for small samples has progressed slower than the corresponding large sample theory. Even so, it is not possible to state that the basic framework of the large sample theory has been completed. Dwass [3] has recently presented a general theorem on the asymptotic normality of certain nonparametric test statistics under alternative hypotheses. His results, however, do not apply to such important and interesting procedures as the c_1-test [11]. Many papers have appeared giving the asymptotic efficiency of particular tests. Hodges and Lehmann [7] have discussed the asymptotic efficiency of the Wilcoxon test with respect to all translation alternatives. In the same paper they have conjectured that the c_1-test is as efficient as the t-test for normal alternatives and at least as efficient as the t-test for all other alternatives.

The beginning of our work came from a desire to verify the Hodges and Lehmann conjecture. Related to the conjecture is the hypothesis that the c_1-statistic is asymptotically normally distributed. Thus our work has two parts: developing a new theorem for asymptotic normality of nonparametric test statistics and the establishing of the variational argument required for determining the minimum efficiency of test procedures.

Received November 26, 1957; revised May 2, 1958.

[1] This work was done with the partial support of the Office of Naval Research while I. R. Savage was at Stanford University and the Center for Advanced Study in the Behavioral Sciences.

Our basic result on the asymptotic normality of statistics of the form T_N is Theorem 1 of Section 4. This theorem is a partial generalization of results of Dwass [3] summarized in our Theorem 4. Theorem 1 is not given in the most general form possible. Our choice of the level of generality was to facilitate our writing and your reading.

Section 3 contains our basic notation and assumptions. Section 4 contains statements of the theorem on asymptotic normality as well as the basic portion of the proof. Details regarding the negligibility of the remainder terms are given in Section 7. The variational arguments are presented in Section 5 and Section 6 relates our Theorem 1 to Dwass's results. Applications of Theorem 1 to several nonparametric tests are given in Section 6.

3. Assumptions and notation. Let $X_1, X_2, \cdots, X_m$ be the ordered observations of a random sample from a population with continuous cumulative distribution function $F(x)$. Let $Y_1, Y_2, \cdots, Y_n$ be the ordered observations of a random sample from a population with continuous cumulative distribution function $G(x)$. Let $N = m + n$ and $\lambda_N = m/N$ and assume that for all N the inequalities $0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$ hold for some fixed $\lambda_0 \leqq \frac{1}{2}$.

Let $F_m(x) =$ (number of $X_i \leqq x)/m$ and $G_n(x) =$ (number of $Y_i \leqq x)/n$. Thus $F_m(x)$ and $G_n(x)$ are the sample cumulative distribution functions of the X's and Y's respectively. Define $H_N(x) = \lambda_N F_m(x) + (1 - \lambda_N) G_n(x)$. Thus $H_N(x)$ is the combined sample cumulative distribution function. The combined population cumulative distribution function is $H(x) = \lambda_N F(x) + (1 - \lambda_N) G(x)$. Even though $H(x)$ depends on N (or rather m and n) through λ_N our notation suppresses this fact for convenience. In fact $F(x)$ and $G(x)$ may actually depend on N although this will not be stated explicitly. In Corollary 1 the distributions do depend on N. The point for suppressing this fact is that our limit theorems are "uniform" and hold, whether the distributions are constant, tend to a limit, or vary rather arbitrarily with the sample size N.

If the ith smallest in the combined sample is an X let $z_{Ni} = 1$ and otherwise let $z_{Ni} = 0$. Then our concern is with statistics of the form

$$mT_N = \sum_{i=1}^{N} E_{Ni} z_{Ni}, \tag{3.1}$$

where the E_{Ni} are given numbers. (The special case where $E_{Ni} = E(i/N)$ is particularly easily handled by our methods. For the Wilcoxon test this condition is met with $E_{Ni} = i/N$, and Freund and Ansari [6] have considered $E_{Ni} = E(i/N) = | \frac{1}{2} - i/N |$ in testing for the equality of dispersion of two populations.) The definition (3.1) of T_N is the one conventionally used. We shall, however, use the following representation:

$$T_N = \int_{-\infty}^{\infty} J_N[H_N(x)]\, dF_m(x). \tag{3.2}$$

The definitions (3.1) and (3.2) are equivalent when E_{Ni} [illegible]

sentation like (3.2) was used by Blum and Weiss [1, page 243, Eq. 2.4] and R
Mises considered $\int \varphi(x)\, dF_m(x)$ in detail [9].

Throughout our proofs K will be used as a generic constant which may depe
on J_N but it will not depend on $F(x)$, $G(x)$, m, n, N. Statements involving
or O_p will always be uniform in $F(x)$, $G(x)$, and $H(x)$, and λ_N in the interval 0
$\lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$.

While J_N need be defined only at $1/N, 2/N, \cdots, N/N$, we shall find it c
venient to extend its domain of definition to $(0,1]$ by some convention such
letting J_N be constant on $(i/N, (i+1)/N]$.

Let I_N be the interval in which $0 < H_N(x) < 1$. Then I_N is closed on the
at the smallest observation and open on the right at the largest observati
The interval, I_N, has a random location.

4. Asymptotic normality.

THEOREM 1. *If*

(1) $J(H) = \lim_{N\to\infty} J_N(H)$ *exists for* $0 < H < 1$ *and is not constant,*

(2) $\int_{I_N} [J_N(H_N) - J(H_N)]\, dF_m(x) = o_p(N^{-1/2})$,

(3) $J_N(1) = o(\sqrt{N})$,

(4) $|J^{(i)}(H)| = \left|\dfrac{d^iJ}{dH^i}\right| \leqq K[H(1-H)]^{-i-\frac{1}{2}+\delta}$

for $i = 0, 1, 2,$ *and for some* $\delta >$

then, for fixed F, G *and* λ_N,

$$\lim_{N\to\infty} P\left(\frac{T_N - \mu_N}{\sigma_N} \leqq t\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx, \tag{4.1}$$

where

$$\mu_N = \int_{-\infty}^{\infty} J[H(x)]\, dF(x) \tag{4.2}$$

and

$$\begin{aligned} N\sigma_N^2 = 2(1-\lambda_N)\Bigg\{ &\iint_{-\infty<x<y<\infty} G(x)[1-G(y)]J'[H(x)]J'[H(y)]\, dF(x)\, dF(y) \\ &+ \frac{(1-\lambda_N)}{\lambda_N} \iint_{-\infty<x<y<\infty} F(x)[1-F(y)]J'[H(x)]J'[H(y)]\, dG(x)\, dG(y) \end{aligned} \tag{4.3}$$

providing $\sigma_N \neq 0$.

In Eqs. 4.1 and 4.3 we put subscripts on μ and σ to recall that these depe
on F, G and λ_N and are meaningful in the more general case where F, G, and
are not fixed. Corollary 1 will extend Theorem 1 to obtain convergence to norm
ity uniformly with respect to F, G, and λ_N for a broad range of F, G, and λ

To facilitate the proof of Corollary 1, we will regard F, G, and λ_N as variable throughout the proof of Theorem 1 except where it is specified otherwise.

Assumption 1 is likely to be filled whenever one speaks of a sequence of tests. In the special case $E_{Ni} = E(i/N)$ of course $J_N = E = J$ and Assumption 2 will automatically be satisfied. Theorem 2 shows that Assumptions 1, 2 and 3 are often satisfied when the E_{Ni} are the mean values of order statistics. Assumption 4 is the basic condition. The assumption has two functions: it limits the growth of the coefficients E_{Ni} and it supplies certain smoothness properties. Both conditions are essential to our argument. We believe that the theorem is true without the smoothness condition.

PROOF. To begin the proof we rewrite T_N as

$$T_N = \int_{-\infty}^{\infty} J_N(H_N)\, dF_m(x) = \int_{I_N} [J_N(H_N) - J(H_N)]\, dF_m(x)$$
$$+ \int_{I_N} J(H_N)\, dF_m(x) + \int_{H_N=1} J_N(H_N)\, dF_m .$$

In the second integral we write $dF_m = d(F_m - F + F)$, $J(H_N) = J(H) + (H_N - H)J'(H) + [(H_N - H)^2/2]J''[\varphi H_N + (1 - \varphi)H]$, where $0 < \varphi < 1$, and $H = \lambda_N F + (1 - \lambda_N)G$. After multiplying out the expression becomes

$$T_N = A + B_{1N} + B_{2N} + \sum_{i=1}^{6} C_{iN} ,$$

where

$$A = \int_{0<H<1} J(H)\, dF(x), \tag{4.4}$$

$$B_{1N} = \int_{0<H<1} J(H)\, d[F_m(x) - F(x)], \tag{4.5}$$

$$B_{2N} = \int_{0<H<1} (H_N - H)J'(H)\, dF(x), \tag{4.6}$$

$$C_{1N} = \lambda_N \int_{0<H<1} (F_m - F)J'(H)\, d[F_m(x) - F(x)] \tag{4.7}$$

$$C_{2N} = (1 - \lambda_N) \int_{0<H<1} (G_n - G)J'(H)\, d[F_m(x) - F(x)], \tag{4.8}$$

$$C_{3N} = \int_{I_N} \frac{(H_N - H)^2}{2} J''[\varphi H_N + (1 - \varphi)H]\, dF_m(x), \tag{4.9}$$

$$C_{4N} = \int_{H_N=1} [- J(H) - (H_N - H)J'(H)]\, dF_m(x), \tag{4.10}$$

$$C_{5N} = \int_{I_N} [J_N(H_N) - J(H_N)]\, dF_m(x), \tag{4.11}$$

$$C_{6N} = \int_{H_N=1} J_N(H_N)\, dF_m(x). \tag{4.12}$$

The A, B, C terms represent the "constant," "first order random," and "higher order random" portions respectively of T_N. In this section a detailed study of the A and B terms is made and in Section 7 it is shown that the C terms are of higher order.

The "constant" term, $A = \int_{0<H<1} J(H)\, dF(x)$, is finite as a result of Assumption 4 of Theorem 1; see Section 7.A.10. Since A depends on λ_N as well as $F(x)$ and $G(x)$ it need not converge as $N \to \infty$, but it does remain bounded.

Integrating B_{2N} by parts and using the fact that

$$\int_{-\infty}^{\infty} d[F_m(x) - F(x)] = 0,$$

we obtain

$$B_{1N} + B_{2N} = [1 - \lambda_N]\left\{\int_{-\infty}^{\infty} B(x)\, d[F_m(x) - F(x)] - \int_{-\infty}^{\infty} B^*(x)\, d[G_n(x) - G(x)]\right\}, \tag{4.13}$$

where

$$B(x) = \int_{x_0}^{x} J'[H(y)]\, dG(y) \tag{4.14}$$

$$B^*(x) = \int_{x_0}^{x} J'[H(y)]\, dF(y) \tag{4.15}$$

and

$$\lambda_N B^*(x) + (1 - \lambda_N)B(x) = J[H(x)] - J[H(x_0)]$$

with x_0 determined somewhat arbitrarily, say by $H(x_0) = 1/2$.

Thus,

$$B_{1N} + B_{2N} = [1 - \lambda_N]\left\{\frac{1}{m}\sum_{i=1}^{m} [(BX_i) - \mathcal{E}B(X)] - \frac{1}{n}\sum_{i=1}^{n} [B^*(Y_i) - \mathcal{E}B^*(Y)]\right\}, \tag{4.16}$$

where $\mathcal{E}$ represents expectation and X and Y have the F and G distributions respectively.

The two summations involve independent samples of identically distributed random variables. Therefore, if F, G, and λ_N are fixed, $B(X)$ and $B^*(Y)$ are specified random variables and we may apply the central limit theorem to show that $B_{1N} + B_{2N}$ when properly normalized has a Gaussian distribution in the limit. The central limit theorem applies if the variances of $B(X)$ and $B^*(Y)$ are finite and at least one is positive.

First, we shall find a bound on the moments of $B(X)$ and $B^*(Y)$:

$$|\, B(x)\, | = \left| \int_{x_0}^{x} J'[H(y)]\, dG(y) \right| \leqq K[H(x)[1 - H(x)]]^{-\frac{1}{2}+\delta}.$$

Thus for $\delta' > 0$ such that $(2 + \delta')(-\frac{1}{2} + \delta) > -1$,

$$\mathcal{E}\{| B(X) |\}^{2+\delta'} \leqq K \int_{-\infty}^{\infty} [H(x)[1 - H(x)]]^{(-\frac{1}{2}+\delta)(2+\delta')} dF(x)$$

$$\leqq K \int_0^1 [H(1 - H)]^{(-\frac{1}{2}+\delta)(2+\delta')} dH \leqq K,$$

having made use of $dG \leqq (1/\lambda_0)\, dH$. (See Section 7.A.8.)

Similarly, we may bound the $2 + \delta'$ absolute moments of $B^*(Y)$. The asymptotic normality of $B_{1N} + B_{2N}$ follows providing $B(X)$ and $B^*(Y)$ do not both have zero variance.

We compute the variances of $B(X)$ and $B^*(Y)$. These can be expressed in terms of $\int B(x)\, dF(x)$, $\int B^2(x)\, dF(x)$, etc, but we shall use a slightly different approach

$$B(X) - \mathcal{E}B(X) = \int_{-\infty}^{\infty} B(x)\, d[F_1(x) - F(x)]$$

$$= -\int_{-\infty}^{\infty} [F_1(x) - F(x)]J'[H(x)]\, dG(x)$$

has variance

$$\sigma^2_{B(X)} = \mathcal{E}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F_1(x) - F(x)][F_1(y) - F(y)]J'[H(x)]J'[H(y)]\, dG(x)\, dG(y)\right\},$$

and

$$\sigma^2_{B(X)} = 2 \iint_{-\infty<x<y<\infty} F(x)[1 - F(y)]J'[H(x)]J'[H(y)]\, dG(x)\, dG(y), \tag{4.17}$$

if it is permitted to interchange expectation and integral That this may be done follows from Fubini's theorem when it is seen that for $x < y$,

$$\mathcal{E}\{| F_1(x) - F(x) |\, | F_1(y) - F(y) |\} \leqq KF(x)[1 - F(y)]$$

and that the last integral above is finite. (In fact this integral is bounded in the argument dealing with (C_{23N}) in Section 7.B.)

Similarly, the variance of $B^*(Y)$ is given by

$$\sigma^2_{B^*(Y)} = 2 \iint_{-\infty<x<y<\infty} G(x)[1 - G(y)]J'[H(x)]J'[H(y)]\, dF(x)\, dF(y). \tag{4.18}$$

These two variances when combined give the variance result stated in (4.3). We review the status of our proof. In Section 7, the C terms are shown to be "higher order uniformly." The A term is non-random and finite. Finally

$$B_{1N} + B_{2N}$$

is the sum of two independent terms each of which is the average of random variables with mean 0 and finite second moments. Theorem 1 follows.

The proof given can be extended to the case where F, G and λ_N are not fixed. To obtain uniform convergence to normality, we apply a theorem of Esseen

([4], p. 43) which is a generalization of the so-called Berry-Esseen theorem ([8], p. 288)[2]. Since the C terms are uniformly $o_p(1/\sqrt{N})$ it suffices to obtain uniform convergence for $B_{1N} + B_{2N}$. For this it suffices to bound $\rho_{2+\delta'}$ for $B(X)$ and $B^*(Y)$. Since we bounded the absolute $2 + \delta'$ moments, all that is required is to bound the variances of $B(X)$ and $B^*(Y)$ away from 0 and to have m and $n \to \infty$. Thus we have

COROLLARY 1. *If the conditions* 1 *to* 4 *of Theorem* 1 *are satisfied, and* F, G, *and* $\lambda_N (0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1)$ *are restricted to a set for which* $B(X)$ *and* $B^*(Y)$ *have variances bounded away from* 0, *then Eq.* 4.1 *(asymptotic normality) holds uniformly with respect to* F, G, *and* λ_N.

COROLLARY 2. *If conditions* 1 *to* 4 *of Theorem* 1 *are satisfied,* $0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$,

$$F(x) = \Psi(x - \theta_N),$$

$$G(x) = \Psi(x - \varphi_N),$$

where Ψ *has a density* ψ, *then Eq.* 4.1 *holds uniformly with respect to* λ_N, θ_N *and* φ_N *for* $\varphi_N - \theta_N$ *in some neighborhood of* 0. *If* $\varphi_N - \theta_N \to 0$,

$$\lim_{N\to\infty} \frac{\lambda_N N \sigma_N^2}{(1 - \lambda_N)} = 2 \iint_{0<x<y<1} x(1 - y) J'(x) J'(y)\, dx\, dy \tag{4.19}$$

$$= \int_0^1 J^2(x)\, dx - \left[\int_0^1 J(x)\, dx\right]^2.$$

PROOF. It suffices to show that $B(X)$ and $B^*(Y)$ have variances bounded away from zero and to establish Eq. 4.19. Since J is not constant and has a second derivative, there is an interval of u in which $J'(u)$ is bounded away from 0 and in which $J'(u) > 0$ or in which $J'(u) < 0$. There is a corresponding interval of x for which $\Psi(x)$ lies in the u interval and its density $\psi(x)$ is almost everywhere bounded away from 0. For $\varphi_N - \theta_N$ small enough, there is an x interval whose length is bounded away from 0 where the densities $f(x) = \psi(x - \theta_N)$ and $g = \psi(x - \varphi_N)$ are almost everywhere bounded away from 0 and $J'[H(x)]$ is bounded away from zero. It follows that $B(X)$ and $B^*(Y)$ have variances bounded away from zero.

All that remains is to establish Eq. 4.19. The first equality follows directly from Theorem 1 by letting $F(x) = x^*$ and $G(x) \to x^*$. The second equality can be obtained by interpreting the double integral as

$$\iiiint_{0<u<x<y<v<1} J'(x) J'(y)\, du\, dx\, dy\, dv$$

[2] Esseen's theorem states that if $X_1, X_2, \cdots, X_n$ are independent observations from a population with mean 0, variance σ^2, and finite absolute $2 + \delta'$ moment $\beta_{2+\delta'}$, $0 < \delta' \leqq 1$, then $|F^* - \Phi^*| < C(\delta') \left[\frac{\rho_{2+\delta'}}{n^{\delta'/2}} + \frac{\rho_{2+\delta'}^{1/\delta'}}{n^{1/2}}\right]$ where F^* is the cdf. of $\bar{X}$, Φ^* is the approximating normal cdf, C depends only on δ' and $\rho_{2+\delta'} = \beta_{2+\delta'}/\sigma^{2+\delta'}$.

and integrating with respect to y first and x second. It can also be obtained by considering a standard derivation [13] of the asymptotic distribution of T_N when $F = G$ where T_N is regarded as the average of a sample of m from the population of N numbers $J_N(1/N), J_N(2/N), \cdots, J_N(N/N)$.

We remark that normalizing J so that $\int_0^1 J(x)\,dx = 0$ and $\int_0^1 J^2(x)\,dx = 1$ will not affect the efficiency of the test. Furthermore, if J is the inverse of a cdf, the right-hand side of (4.19) is the variance of that distribution.

In applying Theorem 1 the verification of condition 2 may cause some difficulty. The following Theorem 2 gives a simple sufficient condition under which conditions 1, 2, and 3 hold. In particular with the use of Theorem 2 it is simple to verify that the distribution of the c_1-statistic does approach a Gaussion distribution for alternative hypotheses.

THEOREM 2 *If $J_N(i/N)$ is the expectation of the ith order statistic of a sample of size N from a population whose cumulative distribution function is the inverse function of J and*

$$| J^{(i)}(u) | \leqq K[u(1 - u)]^{-i-\frac{1}{2}+\delta}, \qquad i = 0, 1, 2,$$

then

$$\lim_{N\to\infty} J_N(H) = J(H), \qquad 0 < H < 1,$$

$$J_N(1) = o(N^{1/2}),$$

and

$$\int_{I_N} [J_N(H_N) - J(H_N)]\, dF_m(x) = o(N^{-1/2}).$$

(We write o instead of o_p because the random sequence is bounded by a nonrandom sequence which is $o(N^{-1/2})$. In fact $| \int [J_N(H_N) - J(H_N)]\, dF_m(x) | \leqq (1/\lambda) \int | J_N(H_N) - J(H_N) |\, dH_N(x)$ and our proof essentially shows that this latter integral which is non-random and independent of F and G, is $o(N^{-1/2})$.)

PROOF. *It is well known that $J_N(H) \to J(H)$. A proof of the other two results* is given in Section 7.C.

5. Variational argument. We have now established that the limiting distribution of the c_1-statistic is Gaussian. Thus we may proceed with the study of the efficiency of this test procedure. We will examine translation alternatives only Since the power of the c_1-test approaches one when the distributions F and G are held fixed as N approaches infinity we restrict our consideration to the following situation.

There is a distribution function $\Psi(x)$ which does not depend on N and $F(x) = \Psi(x - \theta)$ and $G(x) = \Psi(x - \varphi)$. We test the hypothesis that $\Delta = \theta - \varphi = 0$ vs. "near" alternatives of the form $\Delta = \Delta_N = cN^{-1/2}$. We will also assume that

$$0 < \lim_{N\to\infty} \lambda_N = \lambda < 1.$$

With this framework we are able to use the Pitman criterion (the one considered

by Hodges and Lehmann) for finding efficiencies of test procedures. The following conditions have been established for the c_1-statistic if Ψ has a density and clearly hold for the t-statistic if Ψ has finite second moments. There are functions $a_N(\Delta)$ and $b_N(\Delta)$ such that for Δ in some neighborhood of 0,

$$\mathcal{L}\left(\frac{T_N - a_N(\Delta)}{b_N(\Delta)}\right) \Rightarrow N(0, 1), \tag{5.1}$$

$$\operatorname*{limit}_{N\to\infty} \frac{b_N(\Delta_N)}{b_N(0)} = 1, \tag{5.2}$$

$$E_T = \operatorname*{limit}_{N\to\infty} \left[\frac{[a_N(\Delta_N) - a_N(0)]}{\Delta_N N^{1/2} b_N(0)}\right]^2 \tag{5.3}$$

exists and is independent of c.

The quantity E_T is called the efficacy of the procedure based on the sequence of statistics T_N. Of course E_T depends on Ψ. In comparing two sequences of tests, say T_N and T_N^*, for the same pair of near alternatives the two tests will have the same power only when the corresponding sample sizes, N and N^*, satisfy the following relationship

$$\operatorname*{limit}_{N\to\infty} \frac{N^*}{N} = \frac{E_T}{E_{T^*}} = E_{T,T^*} \tag{5.4}$$

if $E_{T^*} \neq 0$. *E_{T,T^*} is called the asymptotic relative efficiency of T_N with respect to T_N^*.*

Let $E_{c_1,t}(\Psi)$ denote the asymptotic efficiency relative to the t-test of the c_1-test against translation alternatives. Then we have $J = J_0$ the inverse of the normal $N(0, 1)$ cdf Φ and applying Corollary 1 and using derivatives in the expression for E_T, we have

$$E_{c_1,t}(\Psi) = I_{1\Psi}^2/\sigma^2, \tag{5.5}$$

where

$$I_{1\Psi} = \int J_0'[\Psi(x)]\psi^2(x)\, dx \tag{5.6}$$

and σ^2 is the variance of the distribution with cdf Ψ (and density ψ).[3] Normalizing Ψ to have mean 0 and variance 1 does not affect $E_{c_1,t}(\Psi)$ which then becomes equal to $I_{1\Psi}^2$. In this section we shall prove

THEOREM 3. *If Ψ is a cdf with a density and finite second moment, then $E_{c_1,t}(\Psi) \geqq 1$, and $E_{c_1,t}(\Psi) = 1$ only if Ψ is normal.*

PROOF. It suffices to show that the minimum of $I_{1\Psi}$ subject to the restrictions

$$I_{2\Psi} = \int x\psi(x)\, dx = 0$$

[3] If Ψ does not have finite variance σ^2, $E_{c_1,t}$ is not defined but it makes sense to regard it as ∞.

and

$$I_{2\Psi} = \int x^2\psi(x)\,dx = 1$$

is attained only for $\Psi = \Phi$ and that $I_{1\Phi} = 1$.

A density $\psi(x)$ assigns to each x a value of Ψ and a corresponding value of $J_0[\Psi(x)]$. If $\psi(x) = 0$ a.e. on an interval, this interval corresponds to a fixed value of $J_0[\Psi(x)]$. If x is then regarded as a function of J_0, it is multivalued at that value of J_0. Otherwise x is continuous and it is increasing in J_0. Conversely any monotone non-decreasing function x of J_0 determines a corresponding cdf Ψ. We have

$$u = \Phi[J_0(u)],$$

$$J_0'(u) = \frac{1}{\varphi[J_0(u)]},$$

and

$$\varphi(t) = \frac{1}{\sqrt{2\pi}}\,e^{-t^2/2}.$$

Furthermore

(5.7) $$\int_{-\infty}^{x} \psi(t)\,dt = \Psi(x) = \int_{-\infty}^{J_0} \varphi(t)\,dt$$

and

$$\psi(x)\,dx = d\Psi(x) = \varphi(J_0)\,dJ_0\,.$$

Consequently our problem consists of finding a monotone function $x(J_0)$ which minimizes

(5.8) $$I_{1\Psi} = \int \frac{1}{\varphi(J_0)}\,\frac{\varphi(J_0)}{\left(\dfrac{dx}{dJ_0}\right)}\,\varphi(J_0)\,dJ_0 = \int \frac{\varphi(J_0)}{\left(\dfrac{dx}{dJ_0}\right)}\,dJ_0$$

subject to the restrictions and

(5 9) $$I_{2\Psi} = \int x\psi(x)\,dx = \int x\varphi(J_0)\,dJ_0 = 0,$$

(5.10) $$I_{3\Psi} = \int x^2\psi(x)\,dx = \int x^2\varphi(J_0)\,dJ_0 = 1.$$

In the above form it is immediately obvious that if $\Psi = \Phi$, $x = J_0$ and hence $I_{1\Phi} = 1$. This form is also more suitable for our variational approach.

Suppose now that x is replaced by $x^* = cx$. Then I_1, I_2 and I_3 are replaced by $I_1^* = I_1/c$, $I_2^* = cI_2$, and $I_3^* = c^2I_3$. Thus if $I_2 = 0$ and $I_3 < 1$, we can obtain $I_2^* = 0$ and $I_3^* = 1$ with $I_1^* < I_1$. This discussion is relevant to the proof of the following lemma.

LEMMA 1. *The solution of the minimization problem is unique if it exists.*

PROOF. Suppose x_1 and x_2 are distinct functions with non-negative derivatives. Then let $x = (1 - w)x_1 + wx_2$, where $0 \leqq w \leqq 1$. Then, by convexity

$$I_1(w) = \int \frac{\varphi(J_0)}{\left(\frac{dx}{dJ_0}\right)} dJ_0 < (1 - w) \int \frac{\varphi(J_0)}{\left(\frac{dx_1}{dJ_0}\right)} dJ_0 + w \int \frac{\varphi(J_0)}{\left(\frac{dx_2}{dJ_0}\right)} dJ_0,$$

$$I_2(w) = \int x\varphi(J_0)\, dJ_0 = (1 - w) \int x_1\varphi(J_0)\, dJ_0 + w \int x_2\varphi(J_0)\, dJ_0,$$

and

$$I_3(w) = \int x^2\varphi(J_0)\, dJ_0 < (1 - w) \int x_1^2\varphi(J_0)\, dJ_0 + w \int x_2^2\varphi(J_0)\, dJ_0.$$

Hence x_1 and x_2 cannot both be solutions of the minimization problem since otherwise a multiple of $(x_1 + x_2)/2$ would then satisfy the side conditions and yield a smaller I_1.

With this lemma, all that remains is to show that $x = J_0$ is a solution of the problem. To this end we establish a sufficient condition for the solution of the problem as follows. Suppose that x_1 and x_2 are monotone functions satisfying the restrictions where x_2 gives a lower value for I_1 than does x_1. Then using the convexity again, we have

$$I_1'(0) = -\int \frac{\frac{d(x_2 - x_1)}{dJ_0}}{\left(\frac{dx_1}{dJ_0}\right)^2} \varphi(J_0)\, dJ_0 < 0,$$

$$I_2'(0) = \int (x_2 - x_1)\varphi(J_0)\, dJ_0 = 0,$$

and

$$I_3'(0) = 2 \int x_1(x_2 - x_1)\varphi(J_0)\, dJ_0 < 0.$$

Consequently we have

LEMMA 2. *If x_1 satisfies the restrictions and if for each x_2 which does so also there is a $\xi \geqq 0$ such that*

$$I_1'(0) + \xi I_3'(0) \geqq 0,$$

then x_1 is the unique solution of the minimization problem.[4]

[4] This sufficient condition is essentially the usual Euler equation except that with the convexity at our disposal and the monotonicity restriction, it plays the role of a sufficient instead of a necessary condition.

Now

$$I_1'(0) = \frac{-(x_2 - x_1)}{\left(\frac{dx_1}{dJ_0}\right)^2} \varphi(J_0) \Big|_{-\infty}^{\infty} + \int (x_2 - x_1) \left[\frac{\varphi'(J_0)}{\left(\frac{dx_1}{dJ_0}\right)^2} - \frac{\frac{2d^2x_1}{dJ_0^2}\varphi(J_0)}{\left(\frac{dx_1}{dJ_0}\right)^3} \right] dJ_0 .$$

Now let $x_1(J_0) = J_0$. Then

$$I_1'(0) + \xi I_3'(0) = \int (x_2 - x_1)[\varphi'(J_0) + 2\xi J_0 \varphi(J_0)]\, dJ_0 ,$$

which vanishes for $\xi = 1/2$. Applying Lemma 2 establishes our theorem.

If we regarded the c_1-test as one tailor made to compete against the best parametric test for translation when F and G are normal, we may inquire about nonparametric tests designed to compete against the best parametric tests when F and G have some other form.

Suppose F and G are known to be of the form $F_0(x - \theta)$ and $F_0(x - \varphi)$ respectively where F_0 has a twice differentiable density f_0. Then an efficient[5] test statistic for $\Delta = \theta - \varphi = 0$ would be the maximum-likelihood estimate

$$\hat{\Delta} = \hat{\theta} - \hat{\varphi}$$

for which the asymptotic distribution is normal with mean Δ and variance $[N\lambda(1 - \lambda)(\text{inf}_{F_0})]^{-1}$, where

$$\text{(5.11)} \qquad \text{inf}_{F_0} = \int \frac{[f_0'(x)]^2}{f_0(x)}\, dx,$$

providing the above integral exists. The relative efficiency of our nonparametric test based on the test statistic T with a specified normalized[6] J to the $\hat{\Delta}$ test is

$$\text{(5.5a)} \qquad E_{T,\hat{\Delta}}(F_0) = \frac{I_{1F_0}^2}{\text{inf}_{F_0}},$$

where

$$\text{(5.6a)} \qquad I_{1F_0} = \int J'(F_0) f_0^2(x)\, dx.$$

It can be shown that the best J in the sense that it maximizes $E_{T,\hat{\Delta}}(F_0)$ is given by

$$\text{(5.12)} \qquad J(u) = \frac{-f_0'(x)}{f_0(x)} (\text{inf}_{F_0})^{-1/2}$$

[5] There seems to be no clear-cut statement in the literature which would establish the test based on $\hat{\Delta}$ as an efficient test invariant under the same translation of the X, and Y, The authors wish to thank the referee who pointed out the following elegant proof. The efficacy of the $\hat{\theta} - \hat{\varphi}$ test is $\lambda(1 - \lambda)\,\text{inf}_{F_0}$, where inf_{F_0} is the information of F_0. No invariant test of $\Delta = \Delta_N$ vs. $\Delta = 0$ can have greater efficacy than the likelihood ratio test for testing $\Delta = \Delta_N$ vs. $\Delta = 0$ when the densities of X and Y are $f_0(x + (1 - \lambda)\Delta)$ and $f_0(x - \lambda\Delta)$. A standard calculation gives this test efficacy $\lambda(1 - \lambda)\,\text{inf}_{F_0}$. Thus our $\hat{\theta} - \hat{\varphi}$ test is efficient

[6] Let J be normalized so that $\int J(u)\, du = 0$ and $\int J^2(u)\, du = 1$

where $u = F_0(x)$. In fact for this J, we have

$$I_{JF} = -(\inf_F)^{-1/2} \int \left[\frac{f_0''(x)}{f_0(x)} - \frac{[f_0'(x)]^2}{f_0^2(x)}\right] \frac{1}{f_0(x)} f_0'(x)\, dx = (\inf_{F_0})^{1/2}$$

and

$$E_{F,J} = 1.$$

As it is to be expected, if $F_0 = \Phi(N(0, 1))$, the corresponding $J = J_0$, the inverse of Φ. The problem of comparing the nonparametric with the parametric procedures designed for F_0 when F and G are translates of $\Psi \neq F_0$ is hindered by our ignorance of the behavior of the *parametric* procedure when $\Psi \neq F_0$.

6. Orientation and applications.

6.A. *Orientation.* In Fraser's book [5] it is shown that the c_1-test has a limiting normal distribution for normal alternatives. We have now shown this to be the case for all alternatives (if we include the cases where $N\sigma_N^2 = 0$ or $N\sigma_N^2 \to 0$ as degenerate cases). Hoeffding's U-statistics include many nonparametric test statistics and he, Lehmann, and Dwass have shown that U-statistics are asymptotically normal under the alternative hypothesis. The U-statistics do not include all statistics of the form

$$(3.1) \qquad mT_N = \sum_{i=1}^{N} E_{N,i} z_{N,i}.$$

In particular c_1 is not a U-statistic. Dwass's results [3], summarized in Theorem 4, appear to be the only useful results for statistics of the form (3.1) under general alternative hypotheses.

THEOREM 4. *Suppose*

(1) *The conditions of the first paragraph of our Section 3 hold (Dwass has written to us indicating that it is sufficient to have m and n approach ∞);*

(2) *The polynomial*

$$P(t) = \sum_{k=1}^{h} b_k t^k$$

is non-degenerate, i.e.,

$$\max(|b_1|, \cdots, |b_h|) > 0;$$

(3) $(X_1, \cdots, X_m, Y_1, \cdots, Y_n) = (U_1, \cdots, U_N)$ *and R_i is the number of U's less than or equal to U_i,*

$$(4) \qquad a_{Ni} = \begin{cases} a_1 = (n/mN)^{1/2}, & i = 1, \cdots, m, \\ a_2 = -(m/nN)^{1/2}, & i = m+1, \cdots, N; \end{cases}$$

$$(5) \qquad t_N = \sum_{i=1}^{N} a_{Ni} P(R_i/N);$$

then

$$\lim_{N\to\infty} P\left(\frac{t_N - E(t_N)}{\sigma_{t_N}} < s\right) = \int_{-\infty}^{s} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx.$$

First note

$$t_N = \sum_{i=1}^{N} P\left(\frac{i}{N}\right)[a_1 z_{N_i} + a_2(1 - z_{N_i})]$$

$$= \sum_{i=1}^{N} P\left(\frac{i}{N}\right) z_{N_i}\left(\frac{1}{\sqrt{N}}\left[\left(\frac{1-\lambda_N}{\lambda_N}\right)^{1/2} + \left(\frac{\lambda_N}{1-\lambda_N}\right)^{1/2}\right]\right) + a_2 \sum_{i=1}^{N} P\left(\frac{i}{N}\right)$$

$$= \sqrt{N} T_N \left(\left(\frac{1-\lambda_N}{\lambda_N}\right)^{1/2} + \left(\frac{\lambda_N}{1-\lambda_N}\right)^{1/2}\right) + K,$$

where in T_N we have $E_{N_i} = P(i/N)$. Thus there is a non-stochastic linear relationship between t_N and T_N. Hence, from the statistical viewpoint t_N is equivalent to T_N, a statistic of the form (3.1). Now let us compare Dwass's conditions with ours.

(1) Requiring λ_N to be bounded away from 0 and 1 seems to be essential in our Theorem 1.

(2) The condition $E_{N_i} = J_N(i/N) = P(i/N) = \sum_{k=1}^{h} b_k(i/N)^k$ is much stronger than our condition 4 in Theorem 1 in two respects: We only require that $J_N(x)$ have a limit and the limit need not be a polynomial in x. Of particular importance we do not require $J(x)$ to be bounded on $0 < x < 1$. The requirement max $(|b_1|, \cdots, |b_h|) > 0$ is to insure that $E_{N_i} \equiv 0$, a trivial case which causes no difficulty.

6.B. *Applications.*

Example 1: Let $E_{N_i} = \sum_{j=i}^{N} j^{-1}$. Then Savage has proved [10] that T_N has a limiting Gaussian distribution under the hypothesis and is the test statistic for the locally most powerful rank test of $\theta_1 = \theta_2$ against the alternative $\theta_1 \neq \theta_2$ where $F(x) = e^{\theta_1 x}$ and $G(x) = e^{\theta_2 x}$, $-\infty < x \leq 0$ and $F(x) = G(x) = 1$, $x > 0$. In order to verify that T_N has a limiting Gaussian distribution under the alternative hypothesis let us check the conditions of Theorem 1. To do so we note that $J_N(i/N)$ is the expected value of the ith smallest observation of a sample from the exponential distribution and that Theorem 2 is applicable. Hence T_N is asymptotically normal in all cases.

Example 2: Van der Waerden [12] has developed the theory of the test statistic

$$T_N = \int_{-\infty}^{\infty} J\left(\frac{NH_N(x)}{N+1}\right) dF_m(x),$$

where J is the inverse of the normal $N(0, 1)$ cumulative distribution. It can be shown that

$$\int_{-\infty}^{\infty} \left| J\left(\frac{NH_N(x)}{N+1}\right) - J(H_N(x)) \right| dH_N(x) = o\left(\frac{1}{\sqrt{N}}\right).$$

Then conditions 2 and 3 of Theorem 1 are established and the asymptotic nor-

mality and efficiency properties for this statistic are verified to be the same as those of the c_1-statistic.

7. Higher order terms. In proving that the C terms of Theorem 1 are uniformly of higher order the following elementary results are used repeatedly.

7.A. *Elementary results.*

1. $H \geqq \lambda_N F \geqq \lambda_0 F$.
2. $H \geqq (1 - \lambda_N)G \geqq \lambda_0 G$.
3. $1 - F \leqq \dfrac{1 - H}{\lambda_N} \leqq \dfrac{1 - H}{\lambda_0}$.
4. $1 - G \leqq \dfrac{1 - H}{1 - \lambda_N} \leqq \dfrac{1 - H}{\lambda_0}$.
5. $F(1 - F) \leqq \dfrac{H(1 - H)}{\lambda_N^2} \leqq \dfrac{H(1 - H)}{\lambda_0^2}$.
6. $G(1 - G) \leqq \dfrac{H(1 - H)}{\lambda_0^2}$.
7. $dH \geqq \lambda_N \, dF \geqq \lambda_0 \, dF$.
8. $dH \geqq (1 - \lambda_N) \, dG \geqq \lambda_0 \, dG$.
9. Let (a_N, b_N) be the interval $S_{N\epsilon}$, where

$$S_{N\epsilon} = \left\{x : H(1 - H) > \frac{\eta_\epsilon \lambda_0}{N}\right\}. \tag{7.1}$$

Then η_ϵ can be chosen independently of F, G and λ_N so that

$$P\{X_i \, \varepsilon \, S_{N\epsilon}, Y_j \, \varepsilon \, S_{N\epsilon}, i = 1, 2, \cdots, m, j = 1, 2, \cdots, n\} \geqq 1 - \epsilon. \tag{7.2}$$

10. $\int_{-\infty}^{\infty} J(H(x)) \, dF(x)$ is finite.

Proof. Using assumption 4 of Theorem 1 and A.7

$$\left| \int_{-\infty}^{\infty} J(H(x)) \, dF(x) \right| \leqq K \int_0^1 [H(1 - H)]^{-\frac{1}{2}+\delta} \, dH \leqq K \int_0^1 \frac{dH}{[H(1 - H)]^{1/2}} \leqq K. \tag{7.3}$$

7.B. *Detailed consideration of the second order terms of Theorem* 1. We are now ready to show that the C terms are uniformly of higher order. We begin with C_{1N} and prove the following identity:

$$C_{1N} = \lambda_N \int_{-\infty}^{\infty} (F_m - F) J'(H) \, d(F_m(x) - F(x)) = \frac{\lambda_N}{2} \left[\int J'(H) \, d(F_m - F)^2 + \frac{1}{m} \int J'(H) \, dF_m \right]. \tag{7.4}$$

Let R be the set of points of increase of F_m. Then the right-hand side of the identity becomes

$$\frac{\lambda_N}{2}\left[\int_{\bar{R}} J'(H)\, d(F_m - F)^2 + \int_R J'(H)\, d(F_m - F)^2 + \frac{1}{m}\sum_{i=1}^{m} J'(H(X_i))\,\frac{1}{m}\right]$$

$$= \frac{\lambda_N}{2}\left[2\int_{\bar{R}} J'(H)(F_m - F)\, d(F_m - F) + \sum_{i=1}^{m} J'(H(X_i)) \cdot \left[\left(\frac{i}{m} - F(X_i)\right)^2 - \left(\frac{i-1}{m} - F(X_i)\right)^2\right] + \frac{1}{m}\sum_{i=1}^{m} J'(H(X_i))\,\frac{1}{m}\right]$$

$$= \frac{\lambda_N}{2}\left[2\int_{\bar{R}} J'(H)(F_m - F)\, d(F_m - F) + \sum_{i=1}^{m} J'(H(X_i)) \cdot \left[\frac{2}{m}\left[\frac{i}{m} - F(X_i)\right] - \frac{1}{m^2}\right] + \sum_{i=1}^{m} J'(H(X_i))\,\frac{1}{m^2}\right]$$

$$= \lambda_N \int (F_m - F) J'(H)\, d(F_m - F).$$

Using this identity we integrate by parts and obtain

$$C_{1N} = -\frac{\lambda_N}{2}\left(C_{11N} + C_{12N} - C_{13N}\right), \tag{7.5}$$

where

$$C_{11N} = \int_{S_{N\epsilon}} (F_m - F)^2 J''(H)\, dH,$$

$$C_{12N} = \int_{\bar{S}_{N\epsilon}} (F_m - F)^2 J''(H)\, dH,$$

$$C_{13N} = \frac{1}{m}\int J'(H(x))\, dF_m(x)$$

$$= \frac{1}{m^2}\sum_{i=1}^{m} J'(H(X_i)),$$

where $S_{N\epsilon}$ was defined in 7.A.9.

Now let us consider the random variable C_{11N}. We find

$$\mathcal{E}\,|\,C_{11N}\,| = \mathcal{E}\left\{\int_{S_{N\epsilon}} (F_m - F)^2\,|\,J''(H)\,|\, dH\right\} = \int_{S_{N\epsilon}} \frac{F(1-F)}{N\lambda_N}\,|\,J''(H)\,|\, dH.$$

Now using assumption 4 of Theorem 1 and 7.A.5 we obtain

$$\mathcal{E}\,|\,C_{11N}\,| \leq \frac{K}{N}\int_{S_{N\epsilon}} \frac{H(1-H)\, dH}{[H(1-H)]^{1-\delta}}$$

$$\leq \frac{K}{N}\int_{\frac{1}{K\delta}}^{1} \frac{1}{H^{1-\delta}}\, dH$$

$$\leq \frac{K}{N^{1+\delta}}.$$

Now using the Markoff inequality ([2], p. 182),

$$\Pr\left(|C_{11N}| > aN^{-1/2}\right) \leqq \frac{K}{N^{\frac{1}{2}+\delta}} \frac{N^{1/2}}{a} = \frac{K}{aN^{\delta}},$$

where K may depend on ϵ. Now consider C_{12N}.

Let $H_1 = H(a_N)$, $H_2 = H(b_N)$ as in 7.A.9. Then $H_1 = 1 - H_2 < K/N$. With probability greater than $1 - \epsilon$ we have

$$C_{12N} = \int_{\bar{S}_{N\epsilon}} (F_m - F)^2 J''(H)\, dH = \int_0^{H_1} F^2 J''(H)\, dH + \int_{H_2}^1 (1 - F)^2 J''(H)\, dH$$

$$|C_{12N}| \leqq K\left[\int_0^{H_1} \frac{H^2\, dH}{(H(1 - H))^{\frac{5}{2}-\delta}} + \int_{H_2}^1 \frac{(1 - H)^2\, dH}{(H(1 - H))^{\frac{5}{2}-\delta}}\right]$$

$$\leqq K \int_0^{H_1} H^{-\frac{1}{2}+\delta}\, dH \leqq KN^{-\frac{1}{2}-\delta}.$$

Hence $C_{11N} + C_{12N}$ which does not involve ϵ is $o_p(N^{-\frac{1}{2}})$. Now to complete the study of C_{1N} we investigate C_{13N}:

$$|C_{13N}| = \frac{1}{m^2}\left|\sum_{i=1}^m J'[H(X_i)]\right| \leqq \frac{K}{m^2}\sum_{i=1}^m [H(X_i)(1 - H(X_i))]^{-\frac{3}{2}+\delta}.$$

We may assume $\delta < \frac{3}{2}$ or $\delta < \frac{1}{2}$ without loss of generality. Then using 7.A.5

$$|C_{13N}| \leqq \frac{K}{N}\frac{1}{m}\sum_{i=1}^m [F(X_i)[1 - F(X_i)]]^{-\frac{3}{2}+\delta},$$

which is distribution free. By a theorem of Marcinkiewicz ([8], pp. 242–243) if a random variable Y has rth order moment finite $(0 < r < 1)$, then the sum of N independent observations on Y is $o_p(N^{1/r})$. If X has cdf F,

$$[F(X)[1 - F(X)]]^{-\frac{3}{2}+\delta}$$

has a finite moment of order $2/(3 - \delta)$ and hence

$$C_{13N} = o_p\left[\frac{1}{m^2} N^{\frac{3}{2}-\frac{\delta}{2}}\right] = o_p[N^{-\frac{1}{2}}].$$

Consequently $C_{1N} = o_p(N^{-\frac{1}{2}})$.

Next consider

$$C_{2N} = (1 - \lambda_N)\int_{-\infty}^{\infty} (G_n - G)J'(H)\, d[F_m(x) - F(x)]. \tag{7.6}$$

We have

$$C_{2N} = (1 - \lambda_N)(C_{21N} + C_{22N})$$

where

$$C_{21N} = \int_{\bar{S}_{N\epsilon}} (G_n - G)J'(H)\, d[F_m(x) - F(x)],$$

$$C_{22N} = \int_{S_{N\epsilon}} (G_n - G)J'(H)\, d[F_m(x) - F(x)].$$

With probability greater than $1 - \epsilon$, there are no observations in $\bar{S}_{N\epsilon}$ and

$$|C_{21N}| \leqq K \int_{S_{N\epsilon}} H(1 - H)[H(1 - H)]^{-1+\delta} \, dH(x) \leqq K \left(\frac{\eta_\epsilon}{N}\right)^{1+\delta}.$$

Since the two samples are independent and $\mathcal{E}(G_n - G) = 0$, we have

$$\mathcal{E}(C_{22N}) = \mathcal{E}\{\mathcal{E}C_{22N} \mid X_1, X_2, \cdots, X_m\} = 0,$$

$$\mathcal{E}(C_{22N}^2 \mid X_1, X_2, \cdots, X_m) = C_{23N} + C_{24N},$$

$$C_{23N} = \frac{2}{n} \iint\limits_{\substack{x,y \in S_{N\epsilon} \\ x<y}} G(x)[1 - G(y)]J'[H(x)]J'[H(y)] \cdot d[F_m(x) - F(x)] \, d[F_m(y) - F(y)],$$

$$C_{24N} = \frac{1}{nm} \int_{S_{N\epsilon}} G(x)[1 - G(x)]\{J'[H(x)]\}^2 \, dF_m(x),$$

$$\mathcal{E}(C_{23N}) = \frac{-2}{nm} \iint\limits_{\substack{x,y \in S_{N\epsilon} \\ x<y}} G(x)[1 - G(y)]J'[H(x)]J'[H(y)] \, dF(x) \, dF(y)^{7}$$

$$\leqq \frac{K}{N^2} \iint\limits_{x<y} H(x)[1 - H(y)] \mid J'[H(x)]J'[H(y)] \mid dH(x) \, dH(y)$$

$$\leqq \frac{K}{N^2} \iint\limits_{0<x<y<1} x^{-1+\delta}(1 - x)^{-1+\delta} y^{-1+\delta}(1 - y)^{-1+\delta} \, dx \, dy \leqq \frac{K}{N^2},$$

$$\mathcal{E}(C_{24N}) = \frac{1}{nm} \int_{S_{N\epsilon}} G(1 - G)(J'[H])^2 \, dF(x) \leqq \frac{K}{N^2} \int_{S_{N\epsilon}} [H(1 - H)]^{-2+2\delta} \, dH(x)$$

$$\leqq \frac{K\eta_\epsilon^{-1+2\delta}}{N^{1+2\delta}} = o(N^{-1}).$$

Hence

$$\mathcal{E}(C_{22N}^2 \mid X_1, X_2, \cdots, X_m) \leqq K o_p(N^{-1}),$$

where K may depend on ϵ and

$$|C_{22N}| \leqq K o_p(N^{-1/2})$$

since

$$P(C_{22N}^2 > a\mathcal{E}(C_{22N}^2 \mid X_1, \cdots, X_m)) < 1/a.$$

Consequently $C_{2N} = (1 - \lambda_N)(C_{21N} + C_{22N})$ which does not involve ϵ, satisfies

$$C_{2N} = o_p(N^{-1/2}).$$

---

[7] This integrand has already appeared as part of the variance in Eq. (4.3).

Now consider

$$(7.7)\quad C_{3N} = \int_{0<H_N(x)<1} [H_N(x) - H(x)]^2 J''[\varphi H_N(x) + (1-\varphi)H(x)]\, dF_m(x),$$

$$0 < \varphi <$$

With probability greater than $1 - \epsilon$, the range of integration $0 < H_N(x) <$ can be replaced by $S_{N,\epsilon}$ without changing C_{3N}. Since

$$(7.8a)\quad \sup_{H_N>0} \left| \frac{H(x)}{H_N(x)} \right| = O_p(1),$$

and

$$(7.8b)\quad \sup_{H_N<1} \left| \frac{1-H(x)}{1-H_N(x)} \right| = O_p(1),$$

for each $\epsilon > 0$, there is an $\eta_\epsilon^* > 0$ such that with probability greater than $1 -$ we have for $0 < H_N(x) < 1$,

$$(7.9)\quad [\varphi H_N + (1-\varphi)H][1 - (\varphi H_N + (1-\varphi)H)] > \eta_\epsilon^* H(x)[1 - H(x)].$$

Then

$$|C_{3N}| \leqq \int_{S_{N,\epsilon}} [H_N(x) - H(x)]^2 (\eta_\epsilon^*)^{-\frac{3}{2}+\delta} \{H[1-H]\}^{-\frac{3}{2}+\delta}\, dF_m(x) = (\eta_\epsilon^*)^{-\frac{3}{2}+\delta} C_{31N}$$

$$\mathcal{E}(|C_{31N}|) \leqq \frac{1}{N} \int_{S_{N,\epsilon}} \left[\lambda_N F(1-F) + \frac{(1-F)(1-2F)}{N} + (1-\lambda_N)G(1-G)\right][H(1-H)]^{-\frac{3}{2}+\delta}\, dF($$

$$\leqq \frac{K}{N} \int_{S_{N,\epsilon}} [H(1-H)]^{-\frac{3}{2}+\delta}\, dH + \frac{K}{N^2} \int_{S_{N,\epsilon}} [H(1-H)]^{-\frac{3}{2}+\delta}\, dF$$

$$\leqq \frac{K\eta_\epsilon^{-\frac{1}{2}+\delta}}{N^{\frac{1}{2}+\delta}} + \frac{K\eta_\epsilon^{-\frac{1}{2}+\delta}}{N^{\frac{1}{2}+\delta}}.$$

Consequently

$$C_{3N} = o_p(N^{-1/2}).$$

The C_{4N} term vanishes unless the greatest of the $N = m + n$ observations is an X. In that case

$$(7.10)\quad C_{4N} = \frac{1}{m}\{-J[H(X_m)] - [1 - H(X_m)]J'[H(X_m)]\}.$$

Using 7.A.9, however,

$$\frac{1}{m} |J[H(X_m)]| \leqq \frac{[H(X_m)[1 - H(X_m)]]^{-\frac{1}{2}+\delta}}{m} \leqq \frac{(\eta_\epsilon^*)^{-\frac{1}{2}+\delta}}{N^{\frac{1}{2}+\delta}}$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

$$[illegible]$$

with a rather standard argument to obtain for $1 < i \leqq (N-1)$,

(7.16)[8] $$g_{i,N}(u) \leqq \sqrt{\frac{(N-1)^3}{2\pi(i-1)(N-i)}}\, e^{-\frac{v^2}{2}\left[\frac{(N-}{(i-1)(N}\right.}$$

where

(7.17) $$v = (N-1)u - (i-1).$$

For $1 < i \leqq N/2$,

(7.18) $$J_N\left(\frac{i}{N}\right) - J\left(\frac{i}{N}\right) = \int_0^1 \left[J(u) - J\left(\frac{i}{N}\right)\right] g_{i,N}(u)$$
$$= D_{11} + D_{12} + D_{21} + D_{22}$$

where

$$D_{11} = \int_0^{u_1} J(u) g_{i,N}(u)\, du, \qquad D_{12} = \int_{1-u_1}^1 J(u$$

$$D_{21} = -\int_0^{u_1} J\left(\frac{i}{N}\right) g_{i,N}(u)\, du, \qquad D_{22} = -\int_{1-u_1}^1 J$$

$$D_3 = \int_{u_1}^{1-u_1} \left(u - \frac{i}{N}\right) J'\left(\frac{i}{N}\right) g_{i,N}(u)\, du$$

$$D_4 = \frac{1}{2}\int_{u_1}^{1-u_1} \left(u - \frac{i}{N}\right)^2 J''(u^*)$$

].

u^* between u and i/N, and

(7.19) $$g_{i,N}(u) = u^{\alpha} \frac{u^{i-1-\alpha}(}{(i -} \quad N!$$

> 0

where $\alpha = \frac{1}{2} - \delta$ and we
ity. Let Φ be the normal

$$|D_{11}| \leqq \int_0^{u_1} K[\qquad du$$

1

(7.20)

[8] K repres
related to th
to the direc

Since $g_{i,N}(u) \geqq g_{i,N}(1-u)$ for $1 < i \leqq N/2$ and $0 \leqq u \leqq 1/2$, $|D_{12}|$ has the same bound as $|D_{11}|$. Similarly

$$(7.21)\qquad |D_{21}| \leqq K\left(\frac{i}{N}\right)^{-a}\Phi\left[\frac{\left(u_1 - \frac{i-1}{N-1}\right)(N-1)^{3/2}}{\sqrt{(i-1)(N-i)}}\right] \leqq KN^a\Phi\left(\frac{-\sqrt{i}}{K}\right)$$

and $|D_{22}|$ has the same bound too. Since the expectation of the ith order statistic from the uniform distribution is $i/(N+1)$,

$$D_3 = -J'\left(\frac{i}{N}\right)\left\{\int_0^{u_1}\left(u - \frac{i}{N}\right)g_{i,N}(u)\,du + \int_{1-u_1}^{1}\left(u - \frac{i}{N}\right)g_{i,N}(u)\,du + \frac{i}{N(N+1)}\right\}$$

Now

$$h(u) = \left|u - \frac{i}{N}\right| g_{i,N}(u) \leqq Kh(1-u) \qquad \text{for } u < u_1.$$

Hence

$$(7.22)\qquad \begin{aligned} |D_3| &\leqq K\left(\frac{i}{N}\right)^{-a-1}\left[K\frac{i}{N}\Phi\left(\frac{-\sqrt{i}}{K}\right) + \frac{i}{N(N+1)}\right] \\ &\leqq KN^a\Phi\left(\frac{-\sqrt{i}}{K}\right) + KN^{a-1} \end{aligned}$$

Finally

$$|D_4| \leqq Ku_1^{-1+\delta}\int_0^1\left(u - \frac{i}{N}\right)^2 g_{i,N}(u)\,du,$$

$$|D_4| \leqq Ku_1^{-1+\delta}\left[\frac{i(N-i+1)}{(N+1)^2(N+2)} + \left(\frac{i}{N+1} - \frac{i}{N}\right)^2\right],$$

$$(7.23)\qquad |D_4| \leqq Ku_1^{-1+\delta}\left[K\frac{u_1}{N} + K\frac{u_1^2}{N^2}\right] \leqq \frac{Ku_1^{-1+\delta}}{N} \leqq \frac{KN^a}{i^{1+a}}.$$

Thus, for $1 < i \leqq N/2$,

$$(7.24)\qquad \left|J_N\left[\frac{i}{N}\right] - J\left[\frac{i}{N}\right]\right| \leqq KN^a\left[\Phi\left(\frac{-\sqrt{i}}{K}\right) + \frac{1}{N} + \frac{1}{i^{1+\delta}}\right]$$

and

$$(7.25)\qquad \begin{aligned} &\left|\int_{1 \leqq NF_m \leqq N/2}[J_N(H_N) - J(H_N)]\,dF_m\right| \\ &\leqq \frac{1}{m}\left\{KN^{1-\delta} + \sum_{i=2}^{N/2}KN^a\left[\Phi\left(\frac{-\sqrt{i}}{K}\right) + \frac{1}{N} \right.\right.\\ &\leqq KN^{-1-\delta} \end{aligned}$$

since $\sum_{i=1}^{\infty} \Phi(-\sqrt{i}/K)$ and $\sum_{i=1}^{\infty} i^{-(1+\alpha)}$ converge. By a symmetric argument we can cover the range $N/2 < NF_m \leqq N$ and our theorem follows.

BIBLIOGRAPHY

[1] J. R. Blum and Lionel Weiss, "Consistency of certain two-sample tests," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 242–246.

[2] Harald Cramér, *Mathematical Methods of Statistics*, Princeton, 1946.

[3] Meyer Dwass, "The large sample powers of rank order tests in the two sample problem," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 352–374.

[4] C. G. Esseen, "Fourier analysis of distribution functions," *Acta Math.*, Vol. 77 (1945), pp. 1–125.

[5] D. A. S. Fraser, *Nonparametric Methods of Statistics*, John Wiley and Sons, Inc., 1956.

[6] J. E. Freund and A. R. Ansari, "Two-way rank sum test for variances," Virginia Polytechnic Institute report supported by the Office of Ordnance Research and National Science Foundation, 1957.

[7] J. L. Hodges, Jr., and Erich L. Lehmann, "The efficiency of some nonparametric competitors of the *t*-test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 324–335.

[8] Michel Loève, *Probability Theory*, D. Van Nostrand Co., Inc., 1955.

[9] R. v. Mises, "On the asymptotic distribution of differentiable statistical functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 309–348.

[10] I. R. Savage, "Contributions to the theory of rank order statistics: Two sample case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 590–616.

[11] M. E. Terry, "Some rank order tests which are most powerful against specific parametric alternatives," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 346–366.

[12] B. L. Van der Waerden, "The Computation of the X-distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1, University of California Press, 1956, pp. 207–208.

[13] Abraham Wald and Jacob Wolfowitz, "Statistical tests based on permutations of the observations," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 358–372.

A HIGH DIMENSIONAL TWO SAMPLE SIGNIFICANCE TEST[1]

By A. P. Dempster

Bell Telephone Laboratories[2], Murray Hill, New Jersey

0. Summary. The classical multivariate 2 sample significance test based on Hotelling's T^2 is undefined when the number k of variables exceeds the number of within sample degrees of freedom available for estimation of variances and covariances. Addition of an a priori Euclidean metric to the affine k-space assumed by the classical method leads to an alternative approach to the same problem. A test statistic F which is the ratio of 2 mean square distances is proposed and 3 methods of attaching a significance level to F are described. The third method is considered in detail and leads to a "non-exact" significance test where the null hypothesis distribution of F depends, in approximation, on a single unknown parameter r for which an estimate must be substituted. Approximate distribution theory leads to 2 independent estimates of r based on nearly sufficient statistics and these may be combined to yield a single estimate. A test of F nominally at the 5% level but based on an estimate of r rather than r itself has a true significance level which is a function of r. This function is investigated and shown to be quite near 5%. The sensitivity of the test to a parameter measuring statistical distance between population means is discussed and it is shown that arbitrarily small differences in each individual variable can result in a detectable overall difference provided the number of variables (or, more precisely, r) can be made sufficiently large. This sensitivity discussion has stated implications for the a priori choice of metric in k-space. Finally a geometrical description of the case of large r is presented.

1. Introduction. The statistical problem here treated is that of significance testing for the difference of the means of 2 k-variate populations which may be assumed to have the same structure of variances and covariances, the test being based on a sample from each population with sample sizes denoted by n_1 and n_2. It is intended to provide a method applicable to data where the number k of characteristics measured on each individual is large but where the number of individuals measured may be quite small. The usual method of classical multivariate statistics encounters a mathematical barrier and becomes inapplicable when $k > n_1 + n_2 - 2$, but certainly the need has arisen in applied statistical work for techniques handling small samples of highly described individuals.

The classical method has 2 equivalent formulations in terms of the T^2 statistic of Hotelling [2] or the best linear discriminator of Fisher [3]. For this method the

Received July 8, 1957; revised June 27, 1958.

[1] Most of the material presented here is also contained in the author's PhD thesis [1] at Princeton University. His work in Princeton was supported principally by the National Research Council of Canada.

[2] Now at Harvard University.

space of the k characteristics is thought of as k-dimensional affine space and needs no further structure: the method is invariant over the choice of any k linear combinations of full rank of the given k variables to be used in place of the given variables. The 2 populations are assumed to be probability distributions over affine k-space and the samples constitute $n_1 + n_2$ points of this space. In the formulation of [3] the sample points are projected along a family of parallel $(k - 1)$-dimensional hyperplanes onto any line, the family being chosen so that the one-dimensional Student's t for the 2 samples is maximized. This t_{max} is then used to test the significance of the difference in population means. However, if $k > n_1 + n_2 - 2$, a family of $(k - 1)$-dimensional hyperplanes can be chosen which projects the points into 2 points, one for each sample. Then $t_{max} = \infty$ regardless of the populations and so is useless as a test-statistic. In the formulation of [2] the samples are used to define a Euclidean metric in the affine k-space and the test-statistic is the distance between the 2 sample means in this metric. This metric is based on the variation of the samples about their means, and if the samples are shifted to have a common mean point and $k > n_1 + n_2 - 2$ the variation spans only a subspace of $n_1 + n_2 - 2$ dimensions. Thus it is not surprising that in this case the method of defining the metric breaks down. Furthermore it is heuristically evident that no metric for a whole affine space can be well-defined from variation taking place in a flat subspace. For these reasons we are forced to give up the classical approach with its elegant mathematical property of affine invariance.

The approach of this paper is based on the observation that, whatever metric is chosen for k-space, the distance between sample means is a statistic which may yield evidence of separation of the populations, and, rather than be preoccupied with a choice of optimum metric from the data, we should try to use a metric determined apart from the data and analyze the information yielded through this metric.

For much of the theory the population distributions will be assumed to be (multivariate) normal.

2. The general method. It is assumed that a Euclidean metric has been assigned to the affine k-space of the k characteristics; that is, k independent linear combinations of the given variables have been chosen which define distance along k mutually orthogonal axes of Euclidean k-space. The metric may be thought of as chosen from a priori knowledge (precise or imprecise) of the joint distributions of the k characteristics, in the hope of roughly sphericalizing these distributions. More detailed remarks on the choice of a metric are to be found in section 5.

Suppose that the 2 population distributions have means denoted by $k \times 1$ vectors ν_1 and ν_2 and common $k \times k$ matrix of variances and covariances denoted by Λ. We are seeking evidence that $\nu_0 = \nu_1 - \nu_2$ is different from zero and are naturally led to consider V_0 the vector joining the sample means. V_0 is an unbiased estimate of ν_0. Having a metric at hand we will try to direct a significance test at the detection of a non-zero length of ν_0 and will use the

length of V_0 in estimating this length. Rejection of the null hypothesis $\nu_0 = 0$ will result from evidence that the length of V_0 is significantly greater than zero.

So far this use of V_0 has been justified mostly on heuristic grounds. It makes sense geometrically. If however we assume that the populations are multivariate normal $N(\nu_1, \Lambda)$ and $N(\nu_2, \Lambda)$ a more mathematical reason may be given. Suppose the $n_1 + n_2$ individuals are regarded as defining a set of orthogonal axes in a Euclidean space of $n_1 + n_2$ dimensions. The space may be regarded as "degree of freedom" (d.f.) space and any set of orthogonal axes defines a set of orthogonal d.f. Such a new set of d.f. may be defined as follows: first choose the d.f. measuring the grand mean of the $n_1 + n_2$ individuals, second choose the d.f. measuring the difference between the means of the 2 samples, and third choose any set of $n_1 + n_2 - 2$ d.f. which together with the first 2 form an orthogonal set. This last set represents "within sample" d.f. Their number $n_1 + n_2 - 2$ will henceforth for convenience be denoted by m. The data, which consists of k points in this $(n_1 + n_2)$-space, can be described by a set of $n_1 + n_2$ $k \times 1$ vectors corresponding to the new d.f. Let U_0 be the vector corresponding to the mean difference d.f. and $U_1, U_2, \cdots, U_m$ be the vectors corresponding to the within sample d.f. It can be easily checked that

$$V_0 = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}} U_0,$$

that $U_1, U_2, \cdots, U_m$ have mean 0, and that $U_0, U_1, \cdots, U_m$ are uncorrelated and each have Λ for matrix of variances and covariances. Finally, assuming normality and defining

$$\xi = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \nu_0,$$

it is seen that $U_0, U_1, \cdots, U_m$ are independent, the first being distributed as $N(\xi, \Lambda)$ and the remainder as $N(0, \Lambda)$. With the normality assumption it is clear that $U_0, U_1, \cdots, U_m$ are sufficient for the parameters ν_0 and Λ, for apart from an irrelevant overall translation of both samples the original data can be reconstructed. But since U_0 is the only one of these vectors involving the parameter ν_0 it is natural to choose a property of U_0 alone in testing significance.

Three methods of testing whether or not U_0 is significantly long will be described, but only the third of these will be pursued. The first is the non-parametric randomization test based on the method of Pitman and Welch [4, 5]. For each of the $\binom{n_1 + n_2}{n_1}$ divisions of the $n_1 + n_2$ individuals into 2 groups of n_1 and n_2 there is a corresponding d.f. for group difference and corresponding vector U. Under the null hypothesis that the $n_1 + n_2$ individuals are a sample from one distribution the lengths of all these vectors U have a joint distribution symmetric under permutation of the vectors. Accordingly U_0 is significantly long at level α if the length of U_0 is beyond the $(1 - \alpha)$ point of the sample cumulative distribution of the set of $\binom{n_1 + n_2}{n_1}$ lengths of vectors U. The second method is

the continuous analogue of the first method which comes into play when normal distributions are assumed. Suppose a set of p d.f. are chosen independently at random uniformly with regard to direction in that part of $(n_1 + n_2)$-space orthogonal to the d.f. for the grand mean. The set of p corresponding vectors together with U_0 have, under the null hypothesis of identical normally distributed populations, joint distributions which are again symmetric under permutations so that a significance test may be defined as in the first method. The limiting test as $p \to \infty$ is uniquely defined and may be regarded as the continuous analogue of the Pitman and Welch procedure. For $k = 1$ this amounts to the usual t test, but for general k the distribution associated with the limiting test appears difficult to handle analytically. However the test could be approximated using a suitable p and experimental sampling.

The third method, which is the concern of most of the subsequent discussion, is also based on normal distribution theory. The idea here is to compare the length of U_0 directly against the lengths of $U_1, U_2, \cdots, U_m$, since under the null hypothesis they form a sample of size $m + 1$ from a certain distribution. Define Q_i = squared length of $U_i (i = 0, 1, \cdots, m)$ and

$$F = Q_0 \Big/ \frac{1}{m} \sum_{i=1}^{m} Q_i$$

Then U_0 will be declared significantly long if F is significantly large. If the null hypothesis distribution of F involved no unknown parameters then an exact test could be based on F; since this is not the case a type of "non-exact significance test" will be introduced.

3. Distribution theory. The distributions involved in the non-exact significance test are those of properties of the vectors $U_0, U_1, U_2, \cdots, U_m$, in particular their lengths and angles between pairs of them. We suppose in this section normal distributions and so may deal with a typical vector U distributed as $N(0, \Lambda)$ or a typical sample of such vectors. Under these assumptions Q, the squared length of U, has the distribution of a quadratic form in k normal variables. Since this distribution in precise form involves k parameters, all unknown, we will rely on the well-known [6] approximation which treats Q as distributed as $\mu\chi_r^2$ depending only on 2 unknown parameters μ and r. The parameters μ and r are generally fitted by equating the first 2 moments, and this results in the inequality $r \leqq k$.

This approximation, at least for integral r, corresponds to approximating the distribution of vector U by a spherical normal distribution lying in a flat subspace of dimension r in k-space. Stated more precisely this says that in the metric chosen for k-space there is an orthogonal transformation to coordinates $(y_1, y_2, \cdots, y_k)$ such that the distribution of U is defined by

(i) density $\dfrac{1}{(2\pi)^{r/2}} \exp\left(-\dfrac{1}{2\mu}\sum_{i=1}^{r} y_i^2\right)$ for $y_1, y_2, \cdots y_r$, and

(ii) $y_{r+1}, y_{r+2}, \cdots, y_k$ are zero with probability one.

Having this approximate underlying distribution for U it is possible to define from it approximate distributions for other statistics based on U.

As a first example consider the angle θ between a pair of vectors U and U' independently distributed according to (i) and (ii) above. Due to the spherical symmetry of the distribution in r-space the conditional distribution of θ given U' does not depend on the particular U' so that the distribution of θ is the distribution of the angle between U and any fixed direction e.g.,

$$y_1 = 1, y_2 = y_3 = \cdots = y_r = 0$$

Thus $\cos^2 \theta$ is distributed as

$$y_1^2/(y_1^2 + \cdots [illegible])$$

i.e. $\cos^2 \theta$ has the β distribution [illegible] defined by density

$$\frac{\Gamma\left(\frac{r}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{r-1}{2}\right)} [illegible]$$

This will be used as an approximation to the distribution of $\cos^2 \theta$ [illegible] circumstances where [illegible]

[illegible]

In particular, [illegible]

$\Gamma_1, \Gamma_2, \ldots, \Gamma_n$ [illegible]

angles between [illegible]

these 2 sets of random [illegible]

[illegible]

[illegible]

so that [illegible]

as a statistic not involving μ for the purpose of estimating r. From joint characteristic functions v and $\sum_{i=1}^{m} Q_i$ are seen to be independent. Thus

$$v \cdot \left(\sum_{i=1}^{m} Q_i\right)^m = \prod_{i=1}^{m} Q_i$$

where the 2 factors on the left are independent as are the m factors on the right. Since the distributions of $\sum_{i=1}^{m} Q_i$ and Q_i are known this equation makes it possible to immediately write down the moments of v about 0 or the cumulants and characteristic function of $\log v$. In this way we approach the limiting χ^2 distribution of $\log v$ as $r \to \infty$ and show that the power series expansion in terms of $(1/r)$ of the cumulants of the actual and asymptotic distributions agree up to the terms in $(1/r)^2$. This asymptotic distribution is stated in [7] to be remarkably good with agreement of the first 4 cumulants to within 5% when r is as small as 5.

Asymptotic expansions for the cumulants may be derived as follows. Define $t = -\log(m^m v)$, and K_s as meaning sth cumulant. Then for any s

$$K_s(\log v) + m^s K_s\left(\log \sum_{i=1}^{m} Q_i\right) = mK_s(\log Q_i),$$

or

$$K_s(\log v) + m^s K_s(\log \chi^2_{mr}) = mK_s(\log \chi^2_r).$$

From [8] asymptotic formulas for the cumulants of $\log \chi^2_n$ are given by

$$\begin{aligned} K_1(\log \chi^2_n) &= \log n - \frac{1}{n} - 2\sum_{j=1}^{\infty} \frac{(-4)^{j-1}B_j}{jn^{2j}} \\ &= \log n - \frac{1}{n} - \frac{1}{3n^2} + \frac{1}{15n^4} - \frac{16}{63n^6} + \cdots, \end{aligned}$$

and

$$\begin{aligned} K_s(\log \chi^2_n) &= (-)^s 2^s \left[\frac{(s-2)!}{2n^{s-1}} + \frac{(s-1)!}{2n^s} + \frac{2}{n^s}\sum_{j=1}^{\infty} \frac{(-4)^{j-1}B_j(2j+s-1)!}{(2j)!\,n^{2j-1}}\right] \\ &= (-)^s 2^s \left[\frac{(s-2)!}{2n^{s-1}} + \frac{(s-1)!}{2n^s} + \frac{s!}{6n^{s+1}} + \frac{0}{n^{s+2}} + \cdots\right] \quad \text{for} \quad s \geqq 2, \end{aligned}$$

where B_j are Bernoulli numbers. Thence

$$\begin{aligned} K_1(t) &= -m \log m - K_1(\log v) \\ &= -m \log m - mK_1(\log \chi^2_r) + mK_1(\log \chi^2_{mr}) \\ &= (m-1)\left[\frac{1}{r} + \frac{1 + \frac{1}{m}}{3r^2} - \frac{2\left(1 + \frac{1}{m} + \frac{1}{m^2} + \frac{1}{m^3}\right)}{15r^4} + \cdots\right], \end{aligned}$$

and for $s \geqq 2$

$$\begin{aligned} K_s(t) &= (-)^s K_s(\log v) = (-)^s[mK_s(\log \chi^2_r) - m^s K_s(\log \chi^2_{mr})] \\ &= 2^{s-1}(s-1)!(m-1)\left[\frac{1}{r^s} + \frac{s\left(1 + \frac{1}{m}\right)}{3r^{s+1}} + \frac{0}{r^{s+2}} + \cdots\right] \end{aligned}$$

Since χ^2_{m-1} has cumulants $K_s = 2^{s-1}(s-1)!\,(m-1)$ it is seen that $t \sim (1/r)\chi^2_{m-1}$ with agreement in first terms of the expansions, and

$$t \sim \left(\frac{1}{r} + \frac{1 + \dfrac{1}{m}}{3r^2}\right)\chi^2_{m-1}$$

with agreement in the first 2 terms, for all cumulants.

Thus r may be estimated by $\hat{r}$ defined by

$$t = \left(\frac{1}{\hat{r}} + \frac{1 + \dfrac{1}{m}}{3\hat{r}^2}\right)(m-1)$$

and for r moderately large the distribution χ^2_{m-1} can be used to put confidence limits on r.

Consider next the set of $\frac{1}{2}m(m-1)$ angles among $U_1, U_2, \cdots, U_m$. Set $n = \frac{1}{2}m(m-1)$ and denote by S_i $(i = 1, 2, \cdots, n)$ the squared sines of these angles. Under the approximate model any S_i is considered distributed as $\beta_{(r-1)/2,1/2}$, but as a further consequence of spherical symmetry in r-space it may be noted that any set of angles containing no closed subset is a mutually independently distributed set, and in particular the angles are pairwise independent. Extending this approximation to complete independence the joint density of the S_i becomes

$$\left(\frac{\Gamma\left(\dfrac{r}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\dfrac{r-1}{2}\right)}\right)^n \prod_{i=1}^{n} (S_i)^{(r-3)/2} \prod_{i=1}^{n} (1 - S_i)^{-\frac{1}{2}}$$

so that $\prod_{i=1}^n S_i$ or $\sum_{i=1}^n \log S_i$ appear as equivalent sufficient statistics for r, so contain approximately all the information about r in the directional properties.

This leads to a consideration of $-\log \beta_{(r-1)/2,1/2}$. The density of $1 - \beta_{(r-1)/2,1/2}$ is easily seen to be asymptotically as $r \to \infty$ the density of $1/r\ \chi^2_1$ and since $\beta_{(r-1)/2,1/2} \to 1$ in probability as $r \to \infty$ it follows that

$$\frac{-\log \beta_{(r-1)/2,1/2}}{1 - \beta_{(r-1)/2,1/2}} \to 1$$

in probability as $r \to \infty$ so that $-\log \beta_{(r-1)/2,1/2}$ is also asymptotically distributed as $1/r\ \chi^2_1$.

Direct asymptotic expansions for the cumulants of $-\log \beta_{(r-1)/2,1/2}$ show that, as with statistic t, this last asymptotic distribution can be modified to have agreement in the first 2 terms. For, since $\chi^2_{r-1} = \beta_{(r-1)/2,1/2} \cdot \chi^2_r$ with independence on the right (as may be seen by computing the characteristic functions of the logs of these random variables),

$$\begin{aligned} K_s(-\log \beta_{(r-1)/2,1/2}) &= (-)^s K_s(\log \beta_{(r-1)/2,1/2}) \\ &= (-)^s[K_s(\log \chi^2_{r-1}) - K_s(\log \chi^2_r)] \end{aligned}$$

for all s. Thence

$$K_1(-\log \beta_{(r-1)/2,1/2}) = \log r + \left[-\frac{1}{r} - \frac{1}{3r^2} + \frac{2}{15r^4} + \cdots\right]$$
$$- \log(r-1) - \left[-\frac{1}{r-1} - \frac{1}{3(r-1)^2} + \frac{2}{15(r-1)^4} + \cdots\right]$$
$$= \frac{1}{r} + \frac{3}{2r^2} + \frac{2}{r^3} + \frac{9}{4r^4} + \cdots,$$

and for $s \geqq 2$

$$K_s(-\log \beta_{(r-1)/2,1/2}) = 2^s\left[\frac{(s-2)!}{2(r-1)^{s-1}} + \frac{(s-1)!}{2(r-1)^s} + \frac{s!}{6(r-1)^s} + 0 + \cdots\right]$$
$$- 2\left[\frac{(s-2)!}{2r^{s-1}} + \frac{(s-1)!}{2r^s} + \frac{s!}{6r^{s-1}} + 0 + \cdots\right]$$
$$= 2^{s-1}(s-1)!\left[\frac{1}{r^s} + \frac{3s}{2r^{s+1}} + \frac{s(s+1)}{r^{s+2}} + \cdots\right]$$

so that

$$-\log \beta_{(r-1)/2,1/2} \sim \left(\frac{1}{r} + \frac{3}{2r^2}\right)\chi_1^2$$

with agreement to second terms in the expansions and therefore usable accuracy for quite small r.

Now we may regard

$$-\sum_{i=1}^{n} \log S_i \sim \left(\frac{1}{r} + \frac{3}{2r^2}\right)\chi_n^2$$

and obtain a new estimate of r. Since in approximation the angles were more than pairwise independent the first 2 moments of this last are asymptotically faithful to the approximate model. The remaining moments however will be distorted slightly on account of non-independence in a way which is difficult to investigate.

Finally an estimate of r can be obtained from $t - \sum_{i=1}^{n} \log S_i$ regarded as asymptotically $(1/r)\ \chi^2_{m-1+n}$ or an appropriate refinement for small r.

4. The non-exact significance test. The question is discussed here of what can be had in the way of a significance test based on $F = Q_0/1/m\sum_{i=1}^{m} Q_i$ considered as $F_{r,mr}$ under the null hypothesis where r is unknown but estimated from a statistic w considered distributed as $f(r)\chi_n^2$ independent of F with $f(r)$ equal to $1/r$ or an asymptotically equivalent refinement of $1/r$. The point estimate of r found from the equation $w = f(r)\cdot n$ will be denoted by $\hat{r}$ and the term $100p\%$ confidence point of r will indicate the value of r satisfying $w = f(r)\chi^2_{n(p)}$ where $\chi^2_{n(p)}$ denotes the $100p\%$ point of χ_n^2. Similar notation will be used for percentage points of other distributions.

A statistical test may be termed exact if the distribution of the test statistic under the null hypothesis does not depend on any unknown parameters. If r

were known the statistic F would have this property and the natural test would be to regard F as significant if $F > F_{r,mr(.95)}$. (Assume for this discussion a standard 5% nominal significance level.) Since r is unknown any test based on F must be non-exact and the natural non-exact test appears to be to regard F as significant if $F > F_{\hat{r},m\hat{r}(.95)}$. This test can also be formulated in terms of quantities α and $\hat{\alpha}$. Define α as the significance level of the observed F as a function of the true parameter r, i.e. α satisfies $F = F_{r,mr(1-\alpha)}$. Similarly $\hat{\alpha}$ as a function of the observed statistics F and w can be determined from $F = F_{\hat{r},m\hat{r}(1-\hat{\alpha})}$. The unattainable exact test is that F is significant if $\alpha < .05$; the non-exact test defined is that F is significant if $\hat{\alpha} < .05$.

The non-exact test still has a significance level (or size or probability of type I error) but this is now a function of r. Denoting this function by $\gamma(r)$ we have

$$\begin{aligned}\gamma(r) &= \Pr(\hat{\alpha} < .05)\\ &= \Pr(F > F_{\hat{r},m\hat{r}(.95)})\\ &= \operatorname*{ave}_{w}\{\Pr(F > F_{\hat{r}(w),m\hat{r}(w)(.95)} \mid w)\}\end{aligned}$$

where F is distributed as $F_{r,mr}$. The last version of $\gamma(r)$ indicates how $\gamma(r)$ can be calculated for given r i.e. by averaging a set of fairly well tabled probabilities over a χ^2 distribution. The major interest of this section is to determine the relation between $\gamma(r)$ and the nominal significance level .05.

The distributions of α and $\hat{\alpha}$ can be compared by fixing α and looking at the variability of the corresponding $\hat{\alpha}$. This amounts to conditioning the various random variables by fixing F to produce the desired α, but leaving w unconditioned. For any fixed α, if r is known, percentage points of w can be translated into percentage points of $\hat{r}$ and thence to percentage points of $\hat{\alpha}$. These are denoted $(\hat{\alpha} \mid \alpha)_{(p)}$. Alternatively, for fixed α, r unknown, but w observed, confidence points for r can be translated into confidence points for α and these will also indicate how much $\hat{\alpha}$ varies about α.

Short of actually calculating $\gamma(r)$ for various values of r, m, n and .05, two arguments will be advanced to show that it is near .05. The first argument is to use a table to back up the belief that the disturbance caused by going from α to $\hat{\alpha}$ is not very great relative to the (0, 1) range of α and is well balanced with regard to direction, so that the unconditional distribution of $\hat{\alpha}$ is not much different from the uniform (0, 1) distribution of α. Table 1 shows quartiles of $(\hat{\alpha} \mid \alpha)$ for $m = 10$, $n = 64$, $\alpha = .05$ and .10, and $r = 6$ and ∞. This table indicates that the

TABLE 1

r	α	$(\hat{\alpha} \mid \alpha)_{(.25)}$	$(\hat{\alpha} \mid \alpha)_{(.75)}$
6	.050	.043	.057
	.100	.091	.107
∞	.050	.030	.070
	.100	.072	.125

disturbance in α caused by using $\hat{\alpha}$ is well-balanced near the 5% level and is slight shift towards 0 near 10%. The indication is that $\gamma(r)$ is very near .05.

The second argument involves computing the non-trivial limit $\gamma(\infty)$: $\lim_{r\to\infty}\gamma(r)$. Define

$$(F_{r,mr}-1)^+ = 0 \quad \text{if} \quad F_{r,mr}-1 \leqq 0$$
$$= F_{r,mr}-1 \quad \text{otherwise.}$$

As

$$r\to\infty \; F_{r,mr} \sim N\left(1, \frac{2}{r}\left\{1+\frac{1}{m}\right\}\right)$$

so that

$$(F_{r,mr}-1)^+ \sim \left(N\left(0, \frac{2}{r}\left\{1+\frac{1}{m}\right\}\right)\right)^+$$

or

$$[(F_{r,mr}-1)^+]^2 \sim 0 \quad \text{or} \quad \frac{2}{r}\left(1+\frac{1}{m}\right)\chi_1^2$$

each with probability $\frac{1}{2}$. Similarly if $(1/\hat{r}) = (1/rn)\chi_n^2$ is put in for $1/r$,

$$[(F_{\hat{r},m\hat{r}}-1)^+]^2 \sim 0 \quad \text{or} \quad \frac{2}{r}\left(1+\frac{1}{m}\right)\chi_1^2 = \frac{2}{r}\left(1+\frac{1}{m}\right)\frac{1}{n}\chi_n^2\cdot\chi_1^2$$

each with probability $\frac{1}{2}$ where χ_n^2 and χ_1^2 are independent. From this

$$\gamma(\infty) = \lim_{r\to\infty}\Pr(F_{r,mr} > F_{\hat{r},m\hat{r}(.95)})$$
$$= \lim_{r\to\infty}\Pr([(F_{r,mr}-1)^+]^2 > [(F_{\hat{r},m\hat{r}}-1)^+]^2_{(.95)})$$
$$= \tfrac{1}{2}\Pr\left(\chi_1^2 > \frac{1}{n}[\chi_n^2\cdot\chi_1^2]_{(.90)}\right)$$

Now

$$\text{ave}\{\chi_1^2\} = 1 \quad \text{and} \quad \text{var}\{\chi_1^2\} = 2,$$

and

$$\text{ave}\left\{\frac{1}{n}\chi \ldots\right\} \ldots \quad \text{and} \quad \text{var}\left\{\frac{1}{n}\ldots\right\} = 2+\frac{6}{n},$$

which indicates stron[illegible]

$$\ldots > \left[\chi_1^2 \ldots\right.$$

[illegible] $(\infty) < .05$ so that [illegible] test is [illegible] r gets [illegible]
foregoing tab[illegible] [illegible]ead [illegible] [illegible]e r tha[illegible]
[illegible]t hope t[illegible] [illegible] $\gamma($[illegible]

In any particular case the spread in [illegible] can be [illegible] confidence points of α from confidence points of [illegible]

One feature of this test which might be regarded [illegible] the non-uniqueness of the vectors [illegible] a choice of an orthogonal set of m d.f. [illegible] $n_2 - 2$ dimensions. The symmetry of the [illegible] an orthogonal set of d.f. [illegible] that the [illegible] for any such set, but there is no [illegible] changed by different choices. In [illegible] variant under all choices so that [illegible] the estimation of r that [illegible] of having used almost all the [illegible] plays only a secondary [illegible] be of minor importance.

5. The sensitivity of the test. [illegible]

[illegible]

This may be contrasted with the statistic T^2 usable if $k \leqq m$, which is non-centrally distributed as

$$T^2 = \frac{m\chi_k^2(\tau_1)}{\chi_{m-k+1}^2}$$

with the numerator independent of the denominator [9] which distribution depends only on τ and on no other properties of ξ. The present situation has the undesirable feature that the G_1-metric may have been selected in such a way that the G_1-length of ξ is too small to cause a significant disturbance in Q_0 whereas τ_1 is large enough to cause a significant disturbance in $\chi_k^2(\tau_1)$. An extreme example of this would occur if the populations were not of full rank but lay in separate parallel hyperplanes but still very close in G_1. Here $\tau = \infty$ but $Q_0(\xi)$ could very well be little disturbed. As long as ξ is regarded as having non-random direction and G_1 cannot be chosen to coincide with G_2 there is danger of insensitivity to a large τ arising from this source. On the other hand if it is permissible to assume randomness for ξ then this danger can be controlled on the average, and further discussion proceeds along these lines.

The high-dimensional case is likely to arise in practice when little or nothing is known about the separating power of individual variables. If nothing is supposed known it may be reasonable to think of ξ as random with all directions intuitively equally likely. The only affine choice consistent with this intuitive notion is to make ξ uniformly distributed with respect to G_2-direction, so the first case considered will be where ξ has constant G_2-length τ_1 and is uniformly distributed with respect to G_2-direction independently of the within sample variation.

Under this assumption and normality $U_0(\xi) = \xi + U_0(0)$ where the 2 vectors on the right are independent each with directions distributed uniformly in G_2. Also, due to the G_2-spherical symmetry of the distribution of $U_0(0)$, the G_2-length of $U_0(0)$ is distributed independently of its G_2-direction. It follows that $U_0(\xi)$ has independently distributed G_2-length and G_2-direction, so that in the equation (5.1) the 2 terms on the right are independent. Also the distribution of $R(U_0)$ is independent of ξ. In our standard approximation of $Q_0(0)$ by $\mu\chi_r^2$ by fitting first 2 moments we have $\text{ave}\{Q_0(0)\} = \mu r$ and $\text{ave}\{Q_0^2(0)\} = \mu^2 r(r+2)$. Also $\text{ave}\{P_0(0)\} = \text{ave}\{\chi_k^2\} = k$ and $\text{ave}\{P_0^2(0)\} = k(k+2)$. Thus

$$\text{ave}\{R(U_0)\} = \frac{\text{ave}\{Q_0(0)\}}{\text{ave}\{P_0(0)\}} = \frac{\mu r}{k}$$

and

$$\text{ave}\{R^2(U_0)\} = \frac{\text{ave}\{Q_0^2(0)\}}{\text{ave}\{P_0^2(0)\}} = \frac{\mu^2 r(r+2)}{k(k+2)}$$

so that

$$\text{ave}\{Q_0(\xi)\} = \text{ave}\{R(U_0)\} \cdot \text{ave}\{\chi_k^2(\tau_1)\} = \mu r\left(1 + \frac{\tau_1^2}{k}\right),$$

$$\text{ave}\,\{Q_0^2(\xi)\} = \text{ave}\,\{R^2(U_0)\}\ \text{ave}\,\{\chi_k^4(\tau_1)\}$$

$$= \mu^2 r(r+2)\left(1 + 2\frac{\tau_1^2}{k} + \frac{\tau_1^4}{k(k+2)}\right),$$

and

$$\text{var}\,\{Q_0(\xi)\} = 2\mu^2 r\left(1 + 2\frac{\tau_1^2}{k} + \frac{k-r}{k+2}\frac{\tau_1^4}{k^2}\right).$$

The distribution of $Q_0(\xi)$ is clearly not non-central χ^2 since the variance of the latter does not involve a term in τ_1^4. For practical purposes it would be reasonable to fit a χ^2 shape to this distribution by fitting first 2 moments, i.e. $\lambda\chi_q^2$ where

$$\lambda = \mu\,\frac{1 + 2\dfrac{\tau_1^2}{k} + \dfrac{k-r}{k+2}\dfrac{\tau_1^4}{k^2}}{1 + \dfrac{\tau_1^2}{k}}$$

$$q = r\left(1 + \frac{\dfrac{r+2}{k+2}\dfrac{\tau_1^4}{k^2}}{1 + 2\dfrac{\tau_1^2}{k} + \dfrac{k-r}{k+2}\dfrac{\tau_1^4}{k^2}}\right)$$

Now it is possible to compute approximate "power functions" and "confidence limits" for τ by assuming F for $\tau > 0$ approximately distributed as $\lambda/\mu\ F_{q,mr}$ and by adopting the procedure used with significance testing of replacing r by $\hat{r}$. These "power functions" and "confidence limits" are actually estimates of the true power functions and confidence limits associated with the non-exact test just as $\hat{\alpha}$ was an estimate of α. The deviation of the estimated power from the true power may again be expected to be near zero and balanced about zero. Using confidence points of r confidence points for any particular value of the power function may be found and these will indicate the order of the disturbance caused by replacing r by $\hat{r}$.

For convenience a criterion different from the power function will be used to measure the sensitivity of the test, namely τ_c the value of τ which will produce on the average a barely significant test statistic. Regarding $(1/m)\sum_{i=1}^{m} Q_i$, the denominator of F, as $(\mu/m)\ \chi_{mr}^2$

$$\text{ave}\,\{F\} = \text{ave}\,\{Q_0(\xi)\}\cdot\text{ave}\left\{\frac{m}{\mu}\cdot\chi_{mr}^{-2}\right\}$$

$$= \mu r\left(1 + \frac{\tau_1^2}{k}\right)\cdot\frac{1}{\mu r}\frac{mr}{mr-2}$$

$$= \left(1 - \frac{2}{mr}\right)^{-1}\left(1 + \frac{\tau_1^2}{k}\right)$$

so that τ_c satisfies

$$1 + \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}\frac{\tau_c^2}{k} = \left(1 - \frac{2}{mr}\right)F_{r,mr(.95)}$$

or, since, for large r, $F_{r,mr} \sim N(1, (2/r)[1 + (1/m)])$, τ_c asymptotically satisfie

$$1 + \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \frac{\tau_c^2}{k} = 1 + 1.65 \left(\frac{2}{r}\right)^{\frac{1}{2}} \left(1 + \frac{1}{m}\right)^{\frac{1}{2}}$$

or

$$\tau_c^2 = N \left(\frac{1}{n_1} + \frac{1}{n_2}\right) k r^{-\frac{1}{2}}$$

where $N = 1.65(2)^{\frac{1}{2}}(1 + (1/m))^{\frac{1}{2}} \doteq 2.3$. Note that for a given experiment $r^{-\frac{1}{2}}$ i the only factor in τ_c^2 which depends on G_1. This result is encouraging, for suppose we have a set of variables with equal but possibly small individual separation parameters ρ. If the within sample variation is independent from variable to variable then $\tau^2 = k\rho^2$. Thus if G_1 can be chosen such that

$$r \geqq \frac{N^2}{\rho^4} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^2$$

then separation would show on the average. This implies that regardless of how small ρ is we need only go on adding variables of separation ρ until r has been built up to correct size. Whether it is possible to continue indefinitely adding variables with small separations in a practical case is uncertain, but the example does show how small individual separations can produce something that will show.

If there is some feeling that ξ is not uniformly distributed relative to G_2-direction an alternative would be to suppose it uniform relative to a different metric G_3 with ellipsoid E_3, i.e. when E_3 appears as the unit sphere ξ appears of length σ and uniform with regard to direction independent of $Q_0(0)$. Then a priori knowledge of the separating powers of the variables could be supposed to consist of some information about E_3. Suppose the mean square G_1-length of ξ is $A^2\sigma^2$ where A^2 depends only on E_3, and suppose the mean square G_1-length of the centrally distributed U_0 is $B^2 = \mu r = \text{ave}\{Q_0(0)\}$. Then

$$\text{ave } \{Q_0(\xi)\} = B^2 + A^2\sigma = \mu r \left(1 + \frac{A^2}{B^2}\sigma^2\right)$$

so that σ_c^2 producing significance on the average is given, in the asymptotic case by

$$\sigma_c^2 = N \left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{A^2}{B^2} r^{-\frac{1}{2}}$$

where now the choice of G_1 can influence both B/A and r.

We are now in a position to discuss theoretical issues concerning the original choice of metric G_1. These suggest that for most purposes the aim should be to make G_2 and G_1 coincide as nearly as possible except for a scale factor. The practical question of how well this can be accomplished is not discussed, nor is it crucial for the use of the method. There are 2 issues in the choice of G_1 : sensitivity of the test and safety of the assumptions.

If G_1 is related to G_2 by a scale factor, then the statistic Q is distributed as $\mu\chi_k^2$ i.e., $r = k$ and under normality the approximation to the distribution of Q by $\mu\chi_r^2$ is exact. This is the way in which choice of G_1 can be made to improve the assumptions. It is heuristically evident that a larger value of r results in less likelihood that approximations of this kind will go wrong.

Regarding sensitivity it can be seen that only when E_2 is G_1-spherical is there equal sensitivity to a separation of τ in all directions and so no danger of insensitivity to large τ. Also it has been seen that when the direction of ξ is assumed G_2-uniformly distributed τ_e^2 depends on $r^{-\frac{1}{2}}$ so again there is evidence that maximizing r to k gives greatest sensitivity. However, under the alternative randomness assumption of ξ uniform over ellipsoid E_3 the situation appears more complicated, for the factor in σ_e^2 to be minimized by choice of G_1 is $A^2/(B^2)\, r^{-\frac{1}{2}}$. This suggests that if something is known about the shape of E_3 as well as E_2 then E_1 should be chosen to give more weight to those directions in which E_3 is long relative to E_2 provided this does not too greatly depress r. It is felt that this last suggestion may be occasionally useful but the general rule will be to try to make $r = k$.

6. Asymptotic behavior. In the foregoing are many results asymptotically true as $r \to \infty$ with m fixed. Certainly these are a mathematical convenience. The question of whether indefinitely large r can be practically obtained remains open. Certainly if k can be made arbitrarily large and each of the k variables contains a part independent of the rest then in theory r can be made arbitrarily large because a metric can be chosen such that $r = k$. What is much more in doubt is whether or not variables could be chosen which would give r indefinitely increasing and τ also increasing at a rate such that the sensitivity of our method would continue to improve.

Whether it is practically attainable or just mathematically useful the following geometrical picture of the asymptotic case is illuminating. Consider throughout the approximate model of section 3 and its asymptotic behavior. As $r \to \infty$ the coefficient of variation of Q (i.e. $\mu\chi_r^2$) tends to 0, so that if we back away from the picture at the correct rate as r increases the vectors $U_1, \cdots, U_m$ will appear to all approach in probability the same constant length. Also since $1 - S_i \sim 1/(r)\, \chi_1^2$ each angle between vectors tends in probability to $\pi/2$ so they tend to an orthogonal set of m equal length vectors. Vector U_0 also becomes perpendicular to $U_1, \cdots, U_m$ but its length depends on τ. However if its length should differ from the common limiting lengths of the rest by a factor as great as $(1 + Nr^{-\frac{1}{2}})^{\frac{1}{2}}$ this is roughly what would be called significant, so that asymptotically a significant U_0 could be indistinguishably different from the rest.

An implication of this asymptotic picture is as follows. For small r it would be natural to compare $Q_0(\xi)$ from U_0 more closely with those Q_i from U_i making the smallest angles with U_0, because if U_0 and U_i are close then $R(U_0)$ and $R(U_i)$ are likely to be more nearly the same. T^2 accomplishes this in a neat manner which disappears when $k > m$, but the present method makes no attempt to do

it. The asymptotic picture says that in the limit there is no hope of making such a correction, for if U_0 is nearly at right angles with every U_i then the radii of E_1 and E_2 in the direction of U_0 bear no relation to the radii in the direction of the U_i, i.e. there are too many directions for U_0 to take to hope that it will be near enough to any U_i to make any difference.

7. Acknowledgments. Heartiest thanks are due to Professor John W. Tukey of Princeton University for his generous guidance in this research.

REFERENCES

[1] A. P. Dempster, "The multivariate two sample problem in the degenerate case," unpublished PhD. thesis, Princeton University, 1956.
[2] H. Hotelling, "The generalisation of 'Student's' ratio," *Ann. Math. Stat.*, Vol. 2 (1931), p. 360.
[3] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*," Vol. 7 (1936), p. 179.
[4] E. J. G. Pitman, "Significance tests which may be applied to samples from any population," J. Roy. Stat. Soc., Supplement, Vol. 4 (1937), p. 119–130, 225–232.
[5] B. L. Welch, "On the z-test in randomized blocks and latin squares," *Biometrika*, Vol. 29 (1937), p. 21.
[6] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance *I*," *Ann. Math. Stat.*, Vol. 25 (1954), p. 290.
[7] M. S. Bartlett and M. G. Kendall, "The statistical analysis of variance heterogeneity and the logarithmic transformation," J. Roy. Stat. Soc., Supplement, Vol. 8 (1946), p. 128.
[8] J. Wishart, "The cumulants of the Z and of the logarithmic χ^2 and t distributions," *Biometrika*, Vol. 34 (1947), p. 170.
[9] R. C. Bose and S. N. Roy, "The distribution of Studentized D^2 statistic," *Sankhya*, Vol. 4 (1938), p. 337.

ON THE KOLMOGOROV AND SMIRNOV LIMIT THEOREMS FOR DISCONTINUOUS DISTRIBUTION FUNCTIONS

By Paul Schmid

Swiss Forest Research Institute and Federal Institute of Technology

1. Introduction. Let $X_1, X_2, \ldots, X_N$ be N independent random variables with the same distribution function $F(x)$. $S_N(x)$ is the empirical distribution function, i.e., $S_N(x) = k/N$ if exactly k of the N values X_i are less than or equal to x. It is of theoretical and practical interest to analyze the behavior of the statistics

$$\sup_{-\infty<x<\infty} |S_N(x) - F(x)| \cdot N^{\frac{1}{2}}$$

and

$$\sup_{-\infty<x<\infty} (S_N(x) - F(x)) \cdot N^{\frac{1}{2}}$$

Kolmogorov [12] proved in a famous paper in 1933 that for $\lambda > 0$

I $$\lim_{N\to\infty} P[\sup_{-\infty<x<\infty} |S_N(x) - F(x)| \cdot N^{\frac{1}{2}} < \lambda] = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2\lambda^2 k^2}$$

if $F(x)$ is a continuous distribution function Smirnov [21] obtained a similar result in 1939, when he showed that

II $$\lim_{N\to\infty} P[\sup_{-\infty<x<\infty} (S_N(x) - F(x)) \cdot N^{\frac{1}{2}} < \lambda] = 1 - e^{-2\lambda^2}$$

holds for continuous distribution functions $F(x)$.

Kolmogorov converts in his proof to a generalization of the Central Limit theorem, whereas Smirnov's theorem was a corollary to a more intricate theorem. But the two formulae can be proved by reciprocal methods They have also been proved by Feller [11] and by Doob [10] and Donsker [9]. Feller made use of characteristic functions and Doob employed stochastic processes. Smirnov [22] found in 1944 the first terms of the asymptotic expansion for the probability in II and an exact formula for finite N. Chung [7] and Blackman [5], [6] were successful in finding the asymptotic expansion for the probability in I.

A somewhat more general form of the statistics, namely

$$\sup_{-\infty<x<\infty} |S_N(x) - F(x)| N^{\frac{1}{2}} \cdot \varphi(F(x)),$$

where $\varphi(y)$ is a positive definite weight function, was discussed by Anderson and Darling [1]. They found the limit distributions for some special weight functions,

Received October 11, 1957; revised June 20, 1958.

by means of stochastic processes. Similar results were obtained by Maniya [16] and Malmquist [15]. Rényi [19], in 1953, established the relations

$$\text{III} \quad \lim_{N\to\infty} P\left[\sup_{a\leq F(x)} \left|\frac{S_N(x) - F(x)}{F(x)}\right| \cdot N^{\frac{1}{2}} < \lambda\right] = \frac{4}{\pi}\sum_{k=0}^{\infty} (-1)^k \frac{\exp\left[-\frac{(2k+1)^2\pi^2}{8}\frac{1-a}{a\lambda^2}\right]}{2k+1}$$

and

$$\text{IV} \quad \lim_{N\to\infty} P\left[\sup_{a\leq F(x)} \left(\frac{S_N(x) - F(x)}{F(x)}\right) N^{\frac{1}{2}} < \lambda\right] = \sqrt{\frac{2}{\pi}} \int_0^{\lambda[a/(1-a)]^{\frac{1}{2}}} e^{-t^2/2}\, dt,$$

where $F(x)$ is a continuous distribution function, $a > 0$, $\lambda > 0$.

The statistics treated here are well suited to test if a sample comes from a population with the distribution function $F(x)$. These test functions have the great advantage in that their distributions are independent of the distribution $F(x)$ of the population. Massey [17], Birnbaum [2], and Malmquist [15] investigated the power of the statistics of Kolmogorov and Smirnov. The limit distributions of these statistics have been tabulated by Smirnov [23], and the distribution for finite N by Massey [18], Birnbaum and Tingey [3], [4]. Rényi tabulated his own limit distributions. Hence, today it is practicable to use these statistics.

In this paper Theorems I through IV are extended for the case of discontinuous distribution function's $F(x)$. The probabilities in question converge also in this case, but the limit distributions are no longer independent of $F(x)$. They depend on the values of $F(x)$ at the discontinuity points, but not on the form of the function between the points of discontinuity. Theorems 1 and 2 are proved by a generalization of the method of Kolmogorov. They can also be proved with the help of stochastic processes, as Doob did it for the case of continuous $F(x)$. We bypass representation of this method since it involves techniques similar to those of Anderson and Darling. The proofs of Theorems 3 and 4 follow in part the methods applied by Rényi, but also make use of the generalization by Kolmogorov of the Central Limit theorem. A part of these results has already been published [20].

I should like to thank W. Saxer for suggesting this topic.

2. Extension of the limit theorems of Kolmogorov and Smirnov. Let $F(x)$ be a distribution function continuous for $x \neq x_\nu$, where $F(x_\nu - 0) = f_{2\nu-1}$, $F(x_\nu) = f_{2\nu}$, for $\nu = 1, 2, \cdots, n$, and $f_{2n+1} = 1$. Denote the corresponding empirical distribution function by $S_N(x)$.

THEOREM 1. *If* $\lambda > 0$, *then*

$$(1) \quad \lim_{N\to\infty} P\,[\sup_{-\infty<x<\infty} |S_N(x) - F(x)| < \lambda N^{-\frac{1}{2}}] = \Phi(\lambda),$$

$$(2) \quad \Phi(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2\lambda^2 k^2}\, c \int \cdots \int_{G_k} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{2n} \Lambda_{ij}\, x_i\, x_j\right] dx_1 \cdots dx_{2n},$$

where

$$\Lambda_{jj} = \frac{f_{j+1} - f_{j-1}}{(f_{j+1} - f_j)(f_j - f_{j-1})}, \qquad \Lambda_{jj-1} = \Lambda_{j-1j} = \frac{-1}{f_j - f_{j-1}},$$

$$\Lambda_{ij} = 0, \qquad \text{for} \quad i < j - 1 \quad \text{or} \quad i > j + 1,$$

$$c = (2\pi)^{-n} \prod_{j=1}^{2n+1} (f_j - f_{j-1})^{-\frac{1}{2}}$$

and

$$G_k = \bigcup_{p_1, p_2, \cdots, p_n = -\infty}^{+\infty} \{-\lambda < x_{2\nu-1} + 2\lambda(p_\nu + kf_{2\nu-1}) < \lambda,$$
$$-\lambda < x_{2\nu} + 2\lambda(p_\nu + kf_{2\nu}) < \lambda, \qquad \nu = 1, \cdots, n\}.$$

THEOREM 2. *If* $\lambda > 0$, *then*

$$\lim_{N \to \infty} P[\sup_{-\infty < x < \infty} (S_N(x) - F(x)) < \lambda N^{-\frac{1}{2}}] = \Phi^+(\lambda), \tag{3}$$

$$\lim_{N \to \infty} P[\sup_{-\infty < x < \infty} (F(x) - S_N(x)) < \lambda N^{-\frac{1}{2}}] = \Phi^+(\lambda), \tag{4}$$

$$\Phi^+(\lambda) = \sum_{k=0}^{1} (-1)^k e^{-2\lambda^2 k^2} c \int \cdots \int_{G_k^+} \exp\left[-\frac{1}{2} \sum_{i,j=1}^{2n} \Lambda_{ij} x_i x_j\right] dx_1 \cdots dx_{2n}, \tag{5}$$

where

$$G_k^+ = \bigcup_{p_1, \cdots, p_n = 0}^{1} \{-\infty < (-1)^{p_\nu}(x_{2\nu-1} + 2\lambda k \cdot f_{2\nu-1}) + 2\lambda p_\nu < \lambda,$$
$$-\infty < (-1)^{p_\nu}(x_{2\nu} + 2\lambda k f_{2\nu}) + 2\lambda p_\nu < \lambda, \qquad \nu = 1, \cdots, n\}.$$

For $\lambda \leqq 0$ all limits are 0. The convergence is uniform in λ in all cases.

If the number of jumps of $F(x)$ is countably infinite, a further limit process has to be made in which at first only the highest jumps of $F(x)$ are taken into account. The two limit processes can be interchanged, because $\Phi(\lambda)$ and $\Phi^+(\lambda)$ are continuous functions of the values of $F(x)$ at the points of discontinuity. Hence further difficulties do not arise in this, the most general case. We will prove Theorem 1 for the case of a distribution function for which the inequalities

$$f_{2\nu+1} > f_{2\nu}, \qquad \nu = 0, 1, \cdots, n,$$

are valid. The results must then hold for any distribution function with n jumps, because both sides of (1) depend continuously on the f's.

If the random variable X has the distribution function $F(x)$, then $Y = F(X)$ is also a random variable, the distribution of which has to fulfill

$$P[F(X) \leqq 0] = 0, \qquad P[F(X) \geqq 1] = 0, \qquad P[f_{2\nu-1} \leqq F(X) < f_{2\nu}] = 0$$

and, for $f_{2\nu} \leqq y \leqq f_{2\nu+1}$,

$$P[F(X) \leqq y] = P[X \leqq F^{-1}(y)] = F(F^{-1}(y)) = y.$$

Furthermore, since

$$P[F(X) = f_{2\nu}] = P[X = x_\nu] = f_{2\nu} - f_{2\nu-1},$$

Y will have the distribution function

$$(6) \qquad F^0(y) = \begin{cases} 0, & \text{for } y \leqq 0, \\ y, & \text{for } f_{2\nu} \leqq y \leqq f_{2\nu+1}, \quad \nu = 0, 1, \cdots, n, \\ f_{2\nu-1}, & \text{for } f_{2\nu-1} \leqq y < f_{2\nu}, \quad \nu = 1, 2, \cdots, n, \\ 1, & \text{for } y \geqq 1. \end{cases}$$

Let $S_N^0(y)$ be the empirical distribution function corresponding to $F^0(y)$. Then we have, for $f_{2\nu} \leqq F(x) \leqq f_{2\nu+1}$,

$$S_N^0(F(x)) = \frac{1}{N} (\text{Number of } F(X_i), F(X_i) \leqq F(x))$$

$$= \frac{1}{N} (\text{Number of } X_i, X_i \leqq x) = S_N(x)$$

and $F^0(F(x)) = F(x)$. Hence

$$\sup_{-\infty<x<\infty} | S_N(x) - F(x) | = \sup_{-\infty<x<\infty} | S_N^0(x) - F^0(x) |,$$

because the other values of $F(x)$ cannot be attained. If we denote by I the union of the closed intervals $[f_{2\nu}, f_{2\nu+1}]$, $\nu = 0, 1, \cdots, n$, we obtain

$$(7) \qquad P[\sup_{-\infty<x<\infty} | S_N(x) - F(x) | < \lambda N^{-\frac{1}{2}}] = P[\sup_{x \varepsilon I} | S_N^0(x) - x | < \lambda N^{-\frac{1}{2}}].$$

Denote by M_N the set of integers j such that $j/N \varepsilon I$,

$$(8) \qquad M_N = \{k_0 = 0, 1, \cdots, k_1; k_2, k_2 + 1, \cdots, k_3; \cdots; k_{2n}, k_{2n} + 1, \cdots, k_{2n+1} = N\}.$$

The k_i are defined such that $k_i/N \to f_i$, as $N \to \infty$. We wish to analyze the behavior of

$$(9) \qquad P\left[\max_{j \varepsilon M_N} \left| S_N^0\left(\frac{j}{N}\right) - \frac{j}{N} \right| < \lambda N^{-\frac{1}{2}}\right]$$

when $N \to \infty$.

The event $\mathcal{E}_{ik}$, $k \varepsilon M_N$, happens if simultaneously all inequalities

$$\left| S_N^0\left(\frac{j}{N}\right) - \frac{j}{N} \right| < \lambda N^{-\frac{1}{2}}, \qquad \text{for } j \leqq k, \ j \varepsilon M_N,$$

and the equality

$$S_N^0\left(\frac{k}{N}\right) - \frac{k}{N} = \frac{i}{N}$$

are fulfilled. P_{ik} is the probability of $\mathcal{E}_{ik}$. P_{0N} is equal to the probability in (9).

We can calculate the P_{ik} recursively by means of the initial conditions $P_{00} = 1$, $P_{i0} = 0$ for $i \neq 0$, and the equations

$$
\begin{aligned}
P_{i\,k+1} &= \sum_j P_{jk}\, P[\mathcal{E}_{i\,k+1} \mid \mathcal{E}_{jk}] \\
(10) \qquad &= \sum_{|j| < \lambda N^{\frac{1}{2}}} P_{jk}\, P\left[S_N^0\left(\frac{k+1}{N}\right) - S_N^0\left(\frac{k}{N}\right)\right. \\
&= \left.\frac{i-j+1}{N} \,\middle|\, S_N^0\left(\frac{k}{N}\right) = \frac{k+j}{N}\right],
\end{aligned}
$$

for $k \varepsilon M_N$, $k + 1 \varepsilon M_N$, and

$$
\begin{aligned}
P_{i\,k_{2\nu}} &= \sum_j P_{j\,k_{2\nu-1}}\, P[\mathcal{E}_{i\,k_{2\nu}} \mid \mathcal{E}_{i\,k_{2\nu-1}}] \\
(11) \qquad &= \sum_{|j| < \lambda N^{\frac{1}{2}}} P_{j\,k_{2\nu-1}}\, P\left[S_N^0\left(\frac{k_{2\nu}}{N}\right) - S_N^0\left(\frac{k_{2\nu-1}}{N}\right)\right. \\
&= \left.\frac{i-j+k_{2\nu}-k_{2\nu-1}}{N} \,\middle|\, S_N^0\left(\frac{k_{2\nu-1}}{N}\right) = \frac{j+k_{2\nu-1}}{N}\right],
\end{aligned}
$$

for $\nu = 1, \cdots, n$.

The occurring conditional probabilities give

$$
\begin{aligned}
(12) \qquad & P\left[S_N^0\left(\frac{k+1}{N}\right) - S_N^0\left(\frac{k}{N}\right) = \frac{i-j+1}{N} \,\middle|\, S_N^0\left(\frac{k}{N}\right) = \frac{k+j}{N}\right] \\
&= \binom{N-k-j}{i-j+1}\left(\frac{1}{N-k}\right)^{i-j+1}\left(\frac{N-k-1}{N-k}\right)^{N-k-i-1},
\end{aligned}
$$

for $k \varepsilon M_N$, $k + 1 \varepsilon M_N$, and

$$
\begin{aligned}
(13) \qquad & P\left[S_N^0\left(\frac{k_{2\nu}}{N}\right) - S_N^0\left(\frac{k_{2\nu-1}}{N}\right) = \frac{i-j+k_{2\nu}-k_{2\nu-1}}{N} \,\middle|\, S_N^0\left(\frac{k_{2\nu-1}}{N}\right) = \frac{k_{2\nu-1}+j}{N}\right] \\
&= \binom{N-k_{2\nu-1}-j}{i-j+k_{2\nu}-k_{2\nu-1}}\left(\frac{k_{2\nu}-k_{2\nu-1}}{N-k_{2\nu-1}}\right)^{i-j+k_{2\nu}-k_{2\nu-1}}\left(\frac{N-k_{2\nu}}{N-k_{2\nu-1}}\right)^{N-k_{2\nu}-i},
\end{aligned}
$$

for $\nu = 1, \cdots, n$, according to the laws of the binomial distribution.

The recursion formulae can be simplified if we introduce the new terms

$$
(14) \qquad Q_{ik} = \frac{N^N(N-k-i)!}{N!(N-k)^{N-k-i}e^k}\, P_{ik}.
$$

Now we have

$$
Q_{00} = 1; \qquad Q_{i0} = 0, \quad \text{for } i \neq 0; \qquad Q_{ik} = 0, \quad \text{for } |i| \geqq \lambda N^{\frac{1}{2}};
$$

$$
Q_{i\,k+1} = \sum_{|j| < \lambda N^{\frac{1}{2}}} Q_{jk}\, \frac{1}{(i-j+1)!}\, e^{-1},
$$

for $k \varepsilon M_N$ and $k + 1 \varepsilon M_N$; $|i| < \lambda N^{\frac{1}{2}}$,

$$
(15) \qquad Q_{i\,k_{2\nu}} = \sum_{|j| < \lambda N^{\frac{1}{2}}} Q_{j\,k_{2\nu-1}}\, \frac{(k_{2\nu}-k_{2\nu-1})^{i-j+k_{2\nu}-k_{2\nu-1}}}{(i-j+k_{2\nu}-k_{2\nu-1})!\,e^{k_{2\nu}-k_{2\nu-1}}},
$$

for $\nu = 1, \cdots, n$, and for the probability (9) we obtain

(16) $$P_{0N} = \frac{N!\,e^N}{N^N} Q_{0N}.$$

For finite N, P_{0N} is evaluable, but as $N \to \infty$ the number of necessary recursion steps tends to infinity.

Let Y_j, $j \varepsilon M_N$, be independent random variables with the distributions

(17) $$P\left[Y_j = \frac{i-1}{\lambda N^{\frac{1}{2}}}\right] = \frac{1}{i!\,e}, \qquad i = 0, 1, 2, \cdots; j \neq k_{2\nu},$$

(18) $$P\left[Y_{k_{2\nu}} = \frac{i - k_{2\nu} + k_{2\nu-1}}{\lambda N^{\frac{1}{2}}}\right] = \frac{(k_{2\nu} - k_{2\nu-1})^i}{i!\,e^{k_{2\nu}-k_{2\nu-1}}}, \qquad i = 0, 1, 2, \cdots.$$

Then

$$E(Y_j) = 0,$$

$$E(Y_j^2) = \frac{1}{\lambda^2 N}, \quad j \neq k_{2\nu}; \qquad E(Y_{k_{2\nu}}^2) = \frac{k_{2\nu} - k_{2\nu-1}}{\lambda^2 N},$$

$$E(|Y_j|^3) = \left(1 + \frac{2}{e}\right)\frac{1}{\lambda^3 N^{\frac{3}{2}}}, \quad j \neq k_{2\nu}; \qquad E(|Y_{k_{2\nu}}|^3) \sim \sqrt{\frac{8}{\pi}}\,\frac{(k_{2\nu} - k_{2\nu-1})^{\frac{3}{2}}}{\lambda^3 N^{\frac{3}{2}}}.$$

The event $\mathfrak{D}_{ik}$, $k \varepsilon M_N$, take place if the inequalities

$$\left|\sum_{l \leqq j} Y_l\right| < 1$$

for all $j \leqq k$ and the equality

$$\sum Y_l = \frac{i}{\lambda N^{\frac{1}{2}}}$$

are simultaneously fulfilled. The [illegible]ty of $\mathfrak{D}_{ik}$ is R_{ik} [illegible] $= 1$, $R_{i0} = 0$ for $i \neq 0$.

We can easily verify that the r[illegible] ulae for the [illegible] e same as for the $Q_{i\nu}$ Therefore,

(19) [illegible]

for all [illegible] the probabilit[illegible]

(20) P_{0} [illegible]

We can [illegible] $n + 1$ re[illegible]

1, [illegible]

(21) R_{ik_2} [illegible]

$R_{ik_{2\nu}}$ [illegible]

ese conditional probabilities can be written in the form

$$P[\mathfrak{D}_{ik_{2\nu+1}} \mid \mathfrak{D}_{jk_{2\nu}}] = P\Big[-1 - \frac{j}{\lambda N^{\frac{1}{2}}} < \sum_{r=k_{2\nu}+1}^{l} Y_r < 1 - \frac{j}{\lambda N^{\frac{1}{2}}},$$

2)

$$l = k_{2\nu} + 1, \cdots, k_{2\nu+1};\ \sum_{r=k_{2\nu}+1}^{k_{2\nu+1}} Y_r = \frac{i-j}{\lambda N^{\frac{1}{2}}}\Big],$$

d their limits for $N \to \infty$ can be obtained by the following lemma of Kolmo-rov.

LEMMA, [12]. *Let $Y_{M1}, \cdots, Y_{Mm_M}$ be, for each M, independent random vari-es, whose values are multiples of $\epsilon = \epsilon(M)$, with*

$$E(Y_{Mj}) = 0, \qquad E(Y_{Mj}^2) = 2b_{Mj}, \qquad E(|Y_{Mj}|^3) = d_{Mj}.$$

t a and b be two numbers such that $a < 0$ and $b > 0$. Assume the existence of sitive numbers $A, \cdots, E$, such that, for all M, the inequalities (i) *through* (iv) *fulfilled:*

(i) $A < \sum_{j=1}^{m_M} b_{Mj} < B$,

(ii) $\frac{d_{Mj}}{b_{Mj}} < C\epsilon$, for all j,

(iii) $P[Y_{Mj} = l_{Mj}\epsilon] > D$ and $P[Y_{Mj} = (l_{Mj} + 1)\epsilon] > D$ for all j and suitably chosen l_{Mj},

(iv) $a + E < i_M\epsilon < b - E$.

hen

$$\Big[a < \sum_{k=1}^{j} Y_{Mk} < b, j = 1, 2, \cdots, m_M; \sum_{k=1}^{m_M} Y_{Mk} = i_M\epsilon\Big]$$

$$= \epsilon\left(u\left(0, 0, i_M\epsilon, 2\sum_{k=1}^{m_M} b_{Mk}\right) + \Delta\right),$$

here $u(\sigma, \tau, s, t)$ is Green's function for the heat equation

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial s^2}$$

n the region G,

$$G = \{a < s < b, t > 0\}.$$

If $\epsilon(M) \to 0$, then $\Delta \to 0$.

This lemma can be applied to the random variables $Y_{k_{2\nu}+1}, Y_{k_{2\nu}+2}, \cdots, Y_{k_{2\nu+1}}$. It should be noticed that the variables $Y_{k_{2\nu}}$ do not fulfill condition (iii) and hence must be treated independently. Our recursion formulae are now

$$R_{i0} = 1, \qquad R_{i0} = 0, \qquad i \neq 0,$$

(23) $$R_{ik_{2\nu+1}} = \sum_{|j|<\lambda N^{\frac{1}{2}}} R_{jk_{2\nu}} \frac{1}{\lambda N^{\frac{1}{2}}}\left(u_1\left(0, 0, \frac{i-j}{\lambda N^{\frac{1}{2}}}, \frac{k_{2\nu+1} - k_{2\nu}}{2\lambda^2 N}\right) + \Delta\right), \qquad \nu = 0, \cdots, n,$$

$$R_{ik_{2\nu}} = \sum_{|j|<\lambda N^{\frac{1}{2}}} R_{jk_{2\nu-1}} \frac{(k_{2\nu} - k_{2\nu-1})^{\cdots}}{(i - j + k_{2\nu} - k_{2\nu-1})^{\cdots}}, \qquad \nu,$$

where $u_j(\sigma, \tau; s, t)$ is Green's function for the heat equation in the region G_j,

$$G_j = \left\{-1 - \frac{j}{\lambda N^{\frac{1}{2}}} < s < 1 - \frac{j}{\lambda N^{\frac{1}{2}}}, t > 0\right\},$$

or

$$(24)\quad \begin{aligned} u_j(\sigma, \tau; s, t) &= \frac{1}{2\sqrt{\pi(t-\tau)}} \sum_{l=-\infty}^{+\infty} (-1)^l \\ &\cdot \exp\left[-\frac{\left(s + \frac{j}{\lambda N^{\frac{1}{2}}} - (-1)^l\left(\sigma + \frac{j}{\lambda N^{\frac{1}{2}}}\right) - 2l\right)^2}{4(t-\tau)}\right]. \end{aligned}$$

If N tends to infinity, the Δ's disappear and the sums over j go over into integrals, with the exception of the sum in the first step which consists of only one summand,

$$R_{jk_1} = \frac{1}{\lambda N^{\frac{1}{2}}}\left(u\left(0, 0; \frac{j}{\lambda N^{\frac{1}{2}}}, \frac{k_1}{2\lambda^2 N}\right) + \Delta\right).$$

With this exception all sums tend to finite positive limits. The factor in (20), multiplied by $N^{-\frac{1}{2}}$ also tends to a finite limit, namely

$$N^{-\frac{1}{2}} \frac{N!\,e^N}{N^N} \sim \sqrt{2\pi}.$$

For $r \cdot N^{-\frac{1}{2}} \to x$, we obtain

$$\frac{N^{\frac{1}{2}}(k_{2\nu} - k_{2\nu-1})^{r+k_{2\nu}-k_{2\nu-1}}}{(r + k_{2\nu} - k_{2\nu-1})!\,e^{k_{2\nu}-k_{2\nu-1}}} \sim \frac{1}{\sqrt{2\pi(f_{2\nu} - f_{2\nu-1})}} \exp\left[-\frac{1}{2}\frac{x^2}{f_{2\nu} - f_{2\nu-1}}\right].$$

Finally we have

$$(25)\quad \begin{aligned} &\lim_{N\to\infty} P\left[\max_{j\in M_N}\left|S_N^0\left(\frac{j}{N}\right) - \frac{j}{N}\right| < \lambda N^{-\frac{1}{2}}\right] \\ &= \sum_{j_0, j_1, \cdots, j_n=-\infty}^{+\infty} (-1)^{\sum_{i=0}^n j_i} (2\pi)^{-n} \prod_{j=1}^{2n+1} (f_j - f_{j-1})^{-\frac{1}{2}} \\ &\cdot \int_{-\lambda < x_j < \lambda} \cdots \int \exp\left[-\frac{1}{2}\sum_{\nu=1}^{n} \frac{(x_{2\nu} - x_{2\nu-1})^2}{f_{2\nu} - f_{2\nu-1}}\right. \\ &\qquad \left. - \frac{1}{2}\sum_{\nu=0}^{n} \frac{(x_{2\nu+1} - (-1)^{j_\nu} x_{2\nu} - 2\lambda j_\nu)^2}{f_{2\nu+1} - f_{2\nu}}\right] dx_1 \cdots dx_{2n}, \end{aligned}$$

where x_0 and x_{2n+1} should be replaced by 0. This expression is $\Phi(\lambda)$.

Let us now prove that for those values of λ and sequences of N for which $\lambda N^{\frac{1}{2}}$ are integers,

$$(26)\quad \lim_{N\to\infty} P[\sup_{x\in I} |S_N^0(x) - x| < \lambda N^{-\frac{1}{2}}] = \lim_{N\to\infty} P\left[\max_{j\in M_N}\left|S_N^0\left(\frac{j}{N}\right) - \frac{j}{N}\right| < \lambda N^{-\frac{1}{2}}\right]$$

must be true.

To each $x \varepsilon I$ there exists a $j \varepsilon M_N$ such that either $x = j/N$ or $x = j/N + \epsilon$ with $0 < \epsilon < 1/N$. Set $S_N^0(x) = i/N$. From

$$S_N^0(x) - x = \frac{i - j}{N} - \epsilon \geqq \frac{\lambda N^{1/2}}{N}$$

follows, for $\epsilon > 0$,

$$S_N^0\left(\frac{j+1}{N}\right) - \frac{j+1}{N} \geqq S_N^0(x) - \frac{j+1}{N} = \frac{i - j - 1}{N} \geqq \frac{\lambda N^{1/2}}{N},$$

because the value to the right is a multiple of $1/N$. From

$$S_N^0(x) - x = \frac{i - j}{N} - \epsilon \leqq -\frac{\lambda N^{1/2}}{N}$$

follows analogously

$$S_N^0\left(\frac{j}{N}\right) - \frac{j}{N} \leqq S_N^0(x) - \frac{j}{N} = \frac{i - j}{N} \leqq -\frac{\lambda N^{1/2}}{N}.$$

The second probability in (26) cannot be smaller than the first one and the limit of the second probability depends continuously on the endpoints of the intervals of I. Therefore the two limits have to be equal The convergence must be uniform in λ, since $\Phi(\lambda)$ is a bounded and continuous function. Hence

$$P\,[\sup_{x \varepsilon I} | S_N^0(x) - x | < \lambda N^{-1/2}]$$

tends to $\Phi(\lambda)$ for all λ and all sequences of N. In view of (7), this proves Theorem 1.

Theorem 2 can be proved in a similar manner. We now disregard the absolute value signs in the definition of $\mathcal{E}_{ik}$ and $\mathfrak{D}_{ik}$. The summations in (10), (11), (15), (21) and (23) go from $-\infty$ to $\lambda N^{1/2}$ and the lower boundaries for the partial sums in (22) are omitted. Green's function for the heat equation in the region G_i^+,

$$G_i^+ = \left\{ s < 1 - \frac{j}{\lambda N^{1/2}}, t > 0 \right\},$$

is now

$$u_i^+(\sigma, \tau; s, t) = \frac{1}{2\sqrt{\pi(t - \tau)}} \sum_{l=0}^{1} (-1)^l \cdot \exp\left[\frac{-\left(s + \frac{j}{\lambda N^{1/2}} - (-1)^l \left(\sigma + \frac{j}{\lambda N^{1/2}}\right) - 2l\right)^2}{4(t - \tau)}\right].$$

Hence

$$\lim_{N \to \infty} P\left[\max_{j \varepsilon M_N}\left(S_N^0\left(\frac{j}{N}\right) - \frac{j}{N}\right) < \lambda N^{-1/2}\right]$$

$$= (2\pi)^{-n} \prod_{l=1}^{2n+1} (t_l - t_{l-1})^{-1/2} \sum_{l_0, l_1, \dots, l_n = 0}^{1} (-1)^{\sum_{i=0}^{n} l_i}$$

$$\cdot \int_{x_j<\lambda} \cdots \int \exp\left[-\frac{1}{2} \sum_{\nu=1}^{n} \frac{(x_{2\nu} - x_{2\nu-1})^2}{f_{2\nu} - f_{2\nu-1}} - \frac{1}{2} \sum_{\nu=0}^{n} \frac{(x_{2\nu+1} - (-1)^{j_\nu} x_{2\nu} - 2\lambda j_\nu)^2}{f_{2\nu+1} - f_{2\nu}} \right] dx_1 \cdots dx_{2n},$$

where again x_0 and x_{2n+1} are 0. This proves Theorem 2.

3. Extension of the limit theorems of Rényi. Let $F(x)$ be a continuous function for $x \neq x_\nu$, with $F(x_\nu - 0) = f_{2\nu-1}$ and $F(x_\nu) = f_{2\nu}$, for $\nu = 1, 2, \cdots, n$, and $f_{2n+1} = 1$. Let f_0 be a positive number such that $f_0 \leqq f_1$. If $f_0 > f_1$, then we get the same results except that only the $f_i \geqq f_0$ will appear. Denote the empirical distribution function by $S_N(x)$.

THEOREM 3. *If* $\lambda > 0$, *then*

$$\lim_{N\to\infty} P\left[\sup_{f_0 \leqq F(x)} \left| \frac{S_N(x) - F(x)}{F(x)} \right| < \lambda N^{-\frac{1}{2}} \right] = \Psi(\lambda), \tag{27}$$

$$\Psi(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k d \int \cdots \int_{H_k} \exp\left[-\frac{1}{2} \sum_{i,j=0}^{2n} \Lambda_{ij} x_i x_j \right] dx_0 \cdots dx_{2n}, \tag{28}$$

where

$$\Lambda_{jj} = \frac{(f_{j+1} - f_{j-1}) f_j^2}{(f_{j+1} - f_j)(f_j - f_{j-1})}, \; \Lambda_{j-1\,j} = \Lambda_{jj-1} = \frac{-f_j f_{j-1}}{(f_j - f_{j-1})},$$

$$\Lambda_{ij} = 0, \quad \text{for } i < j - 1 \quad \text{or} \quad i > j + 1,$$

$$d = (2\pi)^{-n-\frac{1}{2}} \prod_{j=0}^{2n} (f_{j+1} - f_j)^{-\frac{1}{2}} (f_{j+1}^{\frac{1}{2}} f_j^{\frac{1}{2}}),$$

and

$$H_k = \bigcup_{p_1, \cdots, p_n = -\infty}^{+\infty} \{ -\lambda < (-1)^k x_0 + 2\lambda k < \lambda;\ -\lambda < (-1)^{p_\nu} x_{2\nu-1} + 2\lambda p_\nu < \lambda, -\lambda < (-1)^{p_\nu} x_{2\nu} + 2\lambda p_\nu < \lambda, \nu = 1, \cdots, n \}.$$

THEOREM 4. *If* $\lambda > 0$, *then*

$$\lim_{N\to\infty} P\left[\sup_{f_0 \leqq F(x)} \frac{S_N(x) - F(x)}{F(x)} < \lambda N^{-\frac{1}{2}} \right] = \Psi^+(\lambda), \tag{29}$$

$$\lim_{N\to\infty} P\left[\sup_{f_0 \leqq F(x)} \frac{F(x) - S_N(x)}{F(x)} < \lambda N^{-\frac{1}{2}} \right] = \Psi^+(\lambda), \tag{30}$$

$$\Psi^+(\lambda) = \sum_{k=0}^{1} (-1)^k d \int_{H_k^+} \cdots \int \exp\left[-\frac{1}{2} \sum_{i,j} \Lambda_{ij} x_i x_j \right]_{dx_0 \cdots dx_{2n}}, \tag{31}$$

where

$$H_k^+ = \bigcup_{p_1, \cdots, p_n = 0}^{1} \{ -\infty < (-1)^k x_0 + 2\lambda k < \lambda; -\infty < (-1)^{p_\nu} x_{2\nu-1} + 2\lambda p_\nu < \lambda, -\infty < (-1)^{p_\nu} x_{2\nu} + 2\lambda p_\nu < \lambda, \nu = 1, \cdots, n \}.$$

The convergence is in both theorems uniform in λ and for $\lambda \leqq 0$ all limits are 0.

These theorems can also be extended for distribution functions with infinitely many points of discontinuity.

We introduce again the random variable $Y = F(X)$ with the distribution function $F^0(x)$ and the set I as the union of the intervals $[f_{2\nu}, f_{2\nu+1}]$, $\nu = 0, 1, \cdots, n$. For any $F(x) \, \epsilon \, I$ we have

$$\frac{S_N(x) - F(x)}{F(x)} = \frac{S_N^0(F(x)) - F^0(F(x))}{F^0(F(x))},$$

and therefore

$$\sup_{F(x) \geqq f_0} \left| \frac{S_N(x) - F(x)}{F(x)} \right| = \sup_{x \epsilon I} \left| \frac{S_N^0(x) - x}{x} \right|.$$

Let $R_N(x)$ be the empirical distribution function of a sample $Z_1, Z_2, \cdots, Z_N$ from a population with the distribution

(32) $$P[Z \leqq x] = x, \qquad 0 \leqq x \leqq 1,$$

then

(33) $$P\left[\sup_{x \epsilon I} \left| \frac{S_N^0(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right] = P\left[\sup_{x \epsilon I} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right],$$

since the distributions of the two populations coincide for $x \, \epsilon \, I$. Thus,

(34) $$P\left[\sup_{F(x) \geqq f_0} \left| \frac{S_N(x) - F(x)}{F(x)} \right| < \lambda N^{-\frac{1}{2}}\right] = P\left[\sup_{x \epsilon I} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right].$$

The set I_ϵ is defined as the union of the intervals $[f_{2\nu} - \epsilon, f_{2\nu+1} + \epsilon]$, $\nu = 0, 1, \cdots, n$, for $\epsilon > 0$. If $|R_N(x) - x| \leqq \epsilon$, then

$$\sup_{R_N(x) \epsilon I} \left| \frac{R_N(x) - x}{x} \right| \leqq \sup_{x \epsilon I_\epsilon} \left| \frac{R_N(x) - x}{x} \right|,$$

since $R_N(x) \, \epsilon \, I$ implies that $x \, \epsilon \, I_\epsilon$. We see that

(35) $$P\left[\sup_{x \epsilon I_\epsilon} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right] \leqq P[|R_N(x) - x| > \epsilon] + P\left[\sup_{R_N(x) \epsilon I} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right].$$

By a similar procedure we have

(36) $$P\left[\sup_{R_N(x) \epsilon I_{2\epsilon}} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right] \leqq P[|R_N(x) - x| > \epsilon] + P\left[\sup_{x \epsilon I_\epsilon} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right].$$

It is sufficient to prove

(37) $$\lim P\left[\sup_{R_N(x) \epsilon I} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda)$$

since the probability

$$P[\,|R_N(x) - x| > \epsilon]$$

tends to 0 as $N \to \infty$. The function $\Psi(\lambda)$ is continuously dependent on the boundaries of I. Therefore, from (35), (36) and (37) we get

$$\lim_{N\to\infty} P\left[\sup_{x\varepsilon I}\left|\frac{R_N(x) - x}{x}\right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda) \tag{38}$$

and from (34) follows the statement of Theorem 3.

We arrange the numbers $Z_1, \cdots, Z_N$ of the sample according to their values and denote by Z_k^* the one for which there are exactly $k - 1$ smaller numbers in the sample. The probability of ties is 0 since (33) is a continuous distribution function. $R_N(x)$ is equal to k/N in $Z_k^* \leqq x \leqq Z_{k+1}^*$. In this interval

$$\sup_{Z_k^* \leqq x < Z_{k+1}^*}\left|\frac{R_N(x) - x}{x}\right| = \max\left\{\left|\frac{k/N}{Z_k^*} - 1\right|, \left|\frac{k/N}{Z_{k+1}^*} - 1\right|\right\}.$$

For $Z_{k+1}^* \geqq k/N$

$$\left|\frac{k/N}{Z_{k+1}^*} - 1\right| \leqq \left|\frac{(k+1)/N}{Z_{k+1}^*} - 1\right| + \frac{1}{f_0 N}$$

since $k/N \geqq f_0$ implies that $Z_{k+1}^* \geqq f_0$. For $Z_{k+1}^* < k/N$

$$\left|\frac{k/N}{Z_{k+1}^*} - 1\right| \leqq \left|\frac{(k+1)/N}{Z_{k+1}^*} - 1\right|.$$

Therefore,

$$\max_{k/N\varepsilon I_{1/N}}\left|\frac{k/N}{Z_k^*} - 1\right| \leqq \sup_{R_N(x)\varepsilon I_{1/N}}\left|\frac{R_N(x) - x}{x}\right| \leqq \max_{k/N\varepsilon I_{2/N}}\left|\frac{k/N}{Z_k^*} - 1\right| + \frac{1}{f_0 N}$$

and (37) is equivalent to

$$\lim_{N\to\infty} P\left[\max_{k/N\varepsilon I_{1/N}}\left|\frac{k/N}{Z_k^*} - 1\right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda). \tag{39}$$

We can write this equation as

$$\lim_{N\to\infty} P\left[\max_{k/N\varepsilon I_{1/N}}\left|\ln\left(\frac{k/N}{Z_k^*}\right)\right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda) \tag{40}$$

or, since $\log n - \sum_{r=1}^{n} 1/r \to c$,

$$\lim_{N\to\infty} P\left[\max_{k/N\varepsilon I_{1/N}}\left|\ln\frac{1}{Z_k^*} - \sum_{l=k}^{n}\frac{1}{l}\right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda). \tag{41}$$

The random variables $\ln(1/Z_k^*)$ are not independent since they fulfill the inequalities

$$\ln\left(\frac{1}{Z_N^*}\right) < \ln\left(\frac{1}{Z_{N-1}^*}\right) < \cdots < \ln\left(\frac{1}{Z_1^*}\right).$$

However, they do form an additive Markov chain (cf [24]), i.e., their differences

$$\ln\left(\frac{1}{Z_{k-1}^*}\right) - \ln\left(\frac{1}{Z_k^*}\right)$$

are mutually independent. The variables

$$U_l = (N + 1 - l)\left(\ln\frac{1}{Z_{N+1-l}^*} - \ln\frac{1}{Z_{N+2-l}^*}\right), \qquad l = 1, \cdots, N,$$

have the distribution

$$P[U_l \leqq x] = 1 - e^{-x}, \qquad 0 \leqq x < \infty.$$

On the other hand we obtain

$$\ln\left(\frac{1}{Z_k^*}\right) = \sum_{l=1}^{N+1-k} \frac{U_l}{N + 1 - l}$$

and

$$\ln\left(\frac{1}{Z_k^*}\right) - \sum_{l=k}^{N} \frac{1}{l} = \sum_{l=1}^{N+1-k} \frac{U_l - 1}{N + 1 - l}.$$

Some moments of the variables

$$V_l = N^{\frac{1}{2}} \frac{U_l - 1}{N + 1 - l}$$

are

$$E(V_l) = 0, \qquad E(V_l^2) = \frac{N}{(N + 1 - l)^2}, \; E(|V_l|^3) = \left(\frac{12}{e} - 2\right)\frac{N^{\frac{3}{2}}}{(N + 1 - l)^3}.$$

Let the set of integers j, for which $(N + 1 - j)/N \in I_{1/N}$, be

$$\{j_0 = 0, 1, \cdots, j_1; j_2, j_2 + 1, \cdots, j_3; \cdots; j_{2n}, j_{2n} + 1, \cdots, j_{2n} + 1\}.$$

The j_i are defined such that $j_i/N \to 1 - f_{2n+1-i}$ as $N \to \infty$, for $i = 0, 1, \cdots, 2n + 1$.

According to well-known rules for conditional distributions, we have

$$P\left[\max_{k/N \in I_{1/N}} \left|\ln\left(\frac{1}{Z_k^*}\right) - \sum_{l=k}^{N}\frac{1}{l}\right| < \lambda N^{-\frac{1}{2}}\right] = P\left[\max_{k/N \in I_{1/N}} \left|\sum_{l=1}^{N+1-k} V_l\right| < \lambda\right]$$

(42)
$$= \int_{\substack{|x_i| < \lambda \\ i=0,\cdots,2n}} \cdots \int \prod_{\nu=0}^{n} d_{x_{2\nu}} P\left[\max_{l=j_{2\nu}+1,\cdots,j_{2\nu+1}} \left|\sum_{i=1}^{l} V_i\right| < \lambda, \right.$$
$$\left. \sum_{i=1}^{j_{2\nu+1}} V_i \leqq x_{2\nu} \,\middle|\, \sum_{i \leqq j_{2\nu}} V_i = x_{2\nu-1}\right]$$
$$\cdot \prod_{\nu=1}^{n} d_{x_{2\nu-1}} P\left[\sum_{i=1}^{j_{2\nu}} V_i \leqq x_{2\nu-1} \,\middle|\, \sum_{i=1}^{j_{2\nu-1}} V_i = x_{2\nu-2}\right],$$

where $x_{-1} = 0$. The limits of the probabilities which occur in this integral can be calculated by a limit theorem for partial sums of random variables.

LEMMA. (*See* [13], [14].) *Let* $Y_{M1}, \cdots, Y_{Mm_M}$ *be* m_M *independent random variables with*

$$E(Y_{Mk}) = 0, \qquad \sum_{1}^{m_M} E(Y_{Mk}^2) = 2t_M.$$

Assume that for all k

$$\frac{E(|Y_{Mk}|^3)}{E(Y_{Mk}^2)} < \mu(M) \tag{43}$$

where $\mu(M) \to 0$ *as* $M \to \infty$, *and let* a, b, ξ, *and* η *be any numbers such that* $a < 0$, $b > 0$, *and* $a \leqq \xi < \eta \leqq b$. *Then*

$$\lim_{M \to \infty} P\left[a < \sum_{1}^{i} Y_{Mi} < b, i = 1, \cdots, m_M; \xi < \sum_{1}^{m_M} Y_{Mi} < \eta\right] = u(0, 0), \tag{44}$$

where $u(s, t)$ *is the solution of the differential equation*

$$\frac{\partial u}{\partial t} = -\frac{\partial^2 u}{\partial s^2} \tag{45}$$

for which the boundary conditions

$$\begin{aligned} u(s, T) &= 0, \qquad a < s < \xi, \qquad \eta < s < b, \\ u(s, T) &= 1, \qquad \xi < s < \eta, \\ u(a, t) &= 0, \qquad 0 < t < T, \\ u(b, t) &= 0, \qquad 0 < t < T, \end{aligned} \tag{46}$$

are fulfilled.

We can apply this lemma to the variables $Y_{i_{2\nu}+1}, Y_{i_{2\nu}+2}, \cdots, Y_{i_{2\nu+1}}$, for $\nu = 0, 1, \cdots, n$, because these variables satisfy (43), with

$$\mu(N) = \frac{3}{f_0 N^{\frac{1}{2}}}.$$

The sum of the second moments

$$2t_M^{(2\nu+1)} = \sum_{i_{2\nu}+1}^{i_{2\nu+1}} \frac{N}{(N+1-k)^2} = N \sum_{N+1-i_{2\nu+1}}^{N} \frac{1}{k^2} - N \sum_{N+1-i_{2\nu}}^{N} \frac{1}{k^2}$$

tends towards

$$2T^{(2\nu+1)} = \frac{1 - f_{2n-2\nu}}{f_{2n-2\nu}} - \frac{1 - f_{2n-2\nu-1}}{f_{2n-2\nu-1}} = \frac{f_{2n-2\nu-1} - f_{2n-2\nu}}{f_{2n-2\nu-1} f_{2n-2\nu}}, \tag{47}$$

and the boundaries for the partial sums are now

$$a = -\lambda - x_{2\nu-1}, \qquad b = \lambda - x_{2\nu-1}, \qquad \xi = -\lambda - x_{2\nu-1}, \qquad \eta = x_{2\nu} - x_{2\nu-1}. \tag{48}$$

where $x_{-1} = 0$. The solution of (45), which satisfies the boundary conditions (46), (48), is

$$(49)\quad u(s,t) = \frac{1}{2\sqrt{\pi(T^{(2\nu+1)} - t)}} \int_{-\lambda - x_{2\nu-1}}^{x_{2\nu} - x_{2\nu-1}} \sum_{j=-\infty}^{+\infty} (-1)^j \cdot \exp\left[-\frac{(s + x_{2\nu-1} - (-1)^j(x + x_{2\nu-1}) - 2\lambda j)^2}{4(T^{(2\nu+1)} - t)}\right] dx.$$

Hence we have

$$(50)\quad \begin{aligned} &\lim_{N\to\infty} P\left[\max_{l=j_{2\nu}+1,\cdots,j_{2\nu+1}} \left|\sum_{i=1}^{l} V_i\right| < \lambda, \sum_{i=1}^{j_{2\nu+1}} V_i \leq x_{2\nu} \,\middle|\, \sum_{i \leq j_{2\nu}} V_i = x_{2\nu-1}\right] \\ &= -\frac{1}{2\sqrt{\pi T^{(2\nu+1)}}} \int_{-\lambda}^{x_{2\nu}} \sum_{j=-\infty}^{+\infty} (-1)^j \exp\left[-\frac{(x - (-1)^j x_{2\nu-1} - 2\lambda j)^2}{4T^{(2\nu+1)}}\right] dx. \end{aligned}$$

On the other hand, we apply the Central Limit theorem to the variables $V_{j_{2\nu-1}+1}$, $V_{j_{2\nu-1}+2}, \cdots, V_{j_{2\nu}}$, obtaining

$$(51)\quad \lim_{N\to\infty} P\left[\sum_{i=1}^{j_{2\nu}} V_i \leq x_{2\nu-1} \,\middle|\, \sum_{i=1}^{j_{2\nu-1}} V_i = x_{2\nu-2}\right] = \frac{1}{2\sqrt{\pi T^{(2\nu)}}} \int_{\infty}^{x_{2\nu-1} - x_{2\nu-2}} e^{-x^2/4T^{(2\nu)}}\, dx,$$

where the $T^{(2\nu)}$ are defined in the same was as the $T^{(2\nu+1)}$ in (47).

In view of (42), (50) and (51) it follows that

$$(52)\quad \begin{aligned} \lim_{N\to\infty} P\left[\max_{k/N \in I_{1/N}} \left|\sum_{l=1}^{N+1-k} V_l\right| < \lambda\right] &= \frac{1}{2^{2n+1}\pi^{n+\frac12} \prod_{j=1}^{2n+1} (T^{(j)})^{\frac12}} \\ &\cdot \sum_{p_0,\cdots,p_n=-\infty}^{+\infty} \int\cdots\int_{|x_i|<\lambda} \exp\left[-\sum_{\nu=0}^{n} \frac{(x_{2\nu} - (-1)^{p_\nu} x_{2\nu-1} - 2\lambda p_\nu)^2}{4T^{(2\nu+1)}} \right. \\ &\left. - \sum_{\nu=0}^{n-1} \frac{(x_{2\nu+1} - x_{2\nu})}{4T^{(2\nu+2)}}\right] dx_0 \cdots dx_{2n}, \end{aligned}$$

where $x_{-1} = 0$. This expression is $\Psi(\lambda)$. This proves (37) and consequently *Theorem 3.*

Theorem 4 can be proved in the same way. In the lemma of Kolmogorov we replace a and ξ by $-\infty$. The solution of the boundary problem is now

$$(53)\quad u(s,t) = \frac{1}{2\sqrt{\pi(T^{(2\nu+1)} - t)}} \int_{\infty}^{x_{2\nu} - x_{2\nu-1}} \sum_{j=0}^{1} (-1)^j \cdot \exp\left[-\frac{(s + x_{2\nu-1} - (-1)^j(x + x_{2\nu-1}) - 2\lambda j)^2}{4(T^{(2\nu+1)} - t)}\right] dx,$$

and we obtain

$$\lim_{N\to\infty} P\left[\max_{l=i_{2\nu}+1,\cdots,i_{2\nu+1}}\left(\sum_{i=1}^{l} V_i\right) < \lambda, \left(\sum_{i=1}^{i_{2\nu+1}} V_i\right) \leqq x_{2\nu} \,\Big|\, \left(\sum_{i\leqq i_{2\nu}} V_i\right) = x_{2\nu-1}\right] \tag{54}$$
$$= \frac{1}{2\sqrt{\pi T^{(2\nu+1)}}} \int_{-\infty}^{x_{2\nu}} \sum_{j=0}^{1} (-1)^j \exp\left[-\frac{(x-(-1)^j x_{2\nu-1} - 2\lambda j)^2}{4T^{(2\nu+1)}}\right] dx.$$

From that Theorem 4 follows.

REFERENCES

[1] T. W. ANDERSON AND D. A. DARLING, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] Z. W. BIRNBAUM, "On the power of a one-sided test of fit for continuous probability functions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 484–489.

[3] Z. W. BIRNBAUM, "Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size," *J. Amer. Math. Soc.*, Vol. 47 (1952), pp. 425–441.

[4] Z. W. BIRNBAUM AND F. H. TINGEY, "One-sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[5] J. BLACKMAN, "An extension of the Kolmogorov distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 513–520.

[6] J. BLACKMAN, "Correction to 'An extension of the Kolmogorov distribution'," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 318–324.

[7] K. L. CHUNG, "An estimate concerning the Kolmogorov limit distribution," *Trans. Amer. Math. Soc.*, Vol. 67 (1949), pp. 36–50.

[8] D. A. DARLING, "The Kolmogorov-Smirnov, Cramér-von Mises tests," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 823–838.

[9] M. D. DONSKER, "Justification and extension of Doobs heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 277–281.

[10] J. L. DOOB, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[11] W. FELLER, "On the Kolmogorov-Smirnov theorems for empirical distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 177–180.

[12] A. N. KOLMOGOROV, "Sulla determinazione empirica di una legge di distribuzione," *Giorn. Ist. Ital. Attuari*, Vol. 4 (1943), pp. 83–91.

[13] A. N. KOLMOGOROV, "Eine Verallgemeinerung des Laplace-Liapounoffschen Satzes," *Bull. Acad. Sci. USSR*, 1931, pp. 956–962.

[14] A. N. KOLMOGOROV, "Ueber die Grenzwertsätze der Wahrscheinlichkeitsrechnung," *Bull. Acad. Sci. USSR*, 1933, pp. 363–372.

[15] S. MALMQUIST, "On certain confidence contours for distribution functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 523–533.

[16] G. M. MANIYA, "Generalization of the criterion of A. N. Kolmogorov," *Doklady Akad. Nauk SSSR*, Vol. 69 (1949), pp. 495–497. (Russian.)

[17] F. J. MASSEY, JR., "A note on the power of a non-parametric test," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 440–443.

[18] F. J. MASSEY, JR., "A note on the estimation of a distribution function by confidence limits," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 116–119.

[19] A. RÉNYI, "On the theory of order statistics," *Acta Math. Acad. Sci. Hung.*, Vol. 4 (1953), pp. 191–231.

[20] P. SCHMID, "Sur les théorèmes asymptotiques de Kolmogorov et Smirnov pour des fonctions de distributions discontinues," *C. R. Acad. Sci. Paris*, Vol. 243 (1956), pp. 349–351

[21] N. V. SMIRNOV, "Sur les écarts de la courbe de distribution empirique," *Mat. Sbornik*, Vol. 6 (1939), pp. 3–26 (Russian, French summary.)

[22] N. V. SMIRNOV, "Approximation of the distribution law of a random variable by empirical data," *Uspyeki Mat Nauk SSSR*, Vol. 10 (1944), pp. 179–206. (Russian).

[23] N. V. SMIRNOV, "Table for estimating the goodness of fit of empirical distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 279–281.

[24] S. S. WILKS, "Order statistics," *Bull. Amer. Math. Soc*, Vol. 54 (1948), pp. 6–50

INCOMPLETE SUFFICIENT STATISTICS AND SIMILAR TESTS[1]

By Robert A. Wijsman[2]

University of California, Berkeley

0. Summary. For a family of exponential densities a method is given, called "D method," for constructing a class of similar tests in the case that the minimal sufficient statistic is boundedly incomplete. This method also provides a proof of a criterion for bounded incompleteness. Under certain conditions the criterion states that a sufficient statistic for a family of exponential densities is boundedly incomplete if the number of components of the statistic is larger than the number of parameters specifying the distribution. Applications are indicated in the Behrens-Fisher problem, and in the problem of testing the ratio of mean to standard deviation in a normal population. In the latter problem it is shown that the D method generates the whole class of similar tests. Some unsolved problems concerning the existence of an optimal similar test are indicated.

1. Introduction. Lehmann and Scheffé [8], [9] have introduced the concept of completeness of a family of measures and have shown the usefulness of this notion both for unbiased estimation and for the construction of similar regions. The latter were introduced by Neyman and Pearson [11] as a means to cope with tests of composite hypotheses. If the hypothesis is composite only because of nuisance parameters, then the requirement of similarity of the test is often a convenient means of restricting the class of tests to be considered. If the hypothesis is composite both of nuisance parameters and because the parameter tested is not completely specified by the hypothesis, then similarity is often required if the test is to be unbiased. For instance, let θ be a real parameter, τ a possibly vector valued nuisance parameter, and let the hypothesis be $H:\theta \leqq \theta_0$, the alternative $\bar{H}:\theta > \theta_0$, for some specified θ_0. Suppose we want the test to be unbiased, then the power function of the test has to be $\leqq \alpha$ for $\theta \leqq \theta_0$ and $\geqq \alpha$ for $\theta > \theta_0$, where α is the level of significance. If, in addition, the power function is continuous, which is usually the case, then we have automatically that its value on the surface $\theta = \theta_0$ equals α, identically in τ. Search for an optimum unbiased test reduces then to the simpler problem of search for an optimum similar test of the hypothesis $H_1:\theta = \theta_0$ against $\bar{H}:\theta > \theta_0$.

In the presence of a sufficient statistic there exists a special class of easily constructible similar regions [10], termed *similar regions of Neyman structure* by Lehmann and Scheffé [8]. They proved that every similar region is of Neyman structure if and only if the family of distributions of the sufficient statistic, as specified by the hypothesis, is boundedly complete [8]. Unfortunately, there

Received August 29, 1957; revised June 5, 1958.

[1] This investigation was supported (in part) by a research grant (No. G-3666) from the National Institutes of Health, Public Health Service.

[2] Now at the University of Illinois, Department of Mathematics.

are important problems in which the latter condition is not fulfilled, in which case the class of all similar regions is larger than the class of similar regions of Neyman structure. An example is the Behrens-Fisher problem (see, for example, [13], in which also references to earlier work can be found). In this problem the similar regions of Neyman structure are of no use, since for any such region the power function is identically constant.

All remarks in the previous paragraph are equally valid if instead of similar rejection regions we consider randomized similar tests. It is clear from the discussion that in each problem of testing a composite hypothesis by means of a similar test it is important to know whether or not the problem admits a boundedly complete sufficient statistic. If not, one would like to have a method of constructing all similar tests. It is the purpose of this paper to provide partial answers to these problems. In section 3 a method termed the "D method," will be given for the construction of a large class of similar tests in the case of a family of exponential densities. In section 5 the D method will be used to derive a criterion for bounded incompleteness in the case of a family of exponential densities. Two examples of the D method are given in section 4; the first example is the Behrens-Fisher problem, the second example is the problem of testing the ratio of mean to standard deviation in a normal population. For the latter problem it is proved in section 6 that every similar test can be constructed by the D method, provided this method is given sufficiently wide scope. Some remarks on the problem of finding an optimal similar test are made in section 7. A preliminary account of the results of sections 3 and 5 appeared in [16].

2. Similar tests and boundedly incomplete sufficient statistics. Let $\mathfrak{X}$ be a space of points x, $\mathfrak{A}$ a σ-field of subsets of $\mathfrak{X}$ (with $\mathfrak{X} \varepsilon \mathfrak{A}$), and $\mathcal{P} = \{P_\theta, \theta \varepsilon \Omega\}$ a family of probability measures on $(\mathfrak{X}, \mathfrak{A})$. Expectation with respect to P_θ will be denoted by E_θ. If $\omega \subset \Omega$ and T is a sufficient statistic for $\mathcal{P}_\omega = \{P_\theta, \theta \varepsilon \omega\}$, we shall also say that T is a sufficient statistic for ω. The range of T is denoted by $\mathfrak{J}$, and is understood to be a Borel subset of a Euclidean space. Let $\mathfrak{B}$ be the σ-field of Borel subsets of $\mathfrak{J}$. We recall the following definitions: A sufficient statistic for ω is called *minimal* if the sufficient sub σ-field which it induces in $\mathfrak{X}$ is "essentially" contained in every sufficient sub σ-field for ω (see Bahadur [2] for a precise definition).[3] A sufficient statistic T for ω is called *complete* for ω if, for every $\mathfrak{B}$-measurable numerical function g

$$(1) \qquad E_\theta g(T) = 0 \quad \text{for all} \quad \theta \varepsilon \omega \Rightarrow g \equiv 0 \quad \text{a.e.} \quad (\mathcal{P}_\omega).$$

If the implication (1) holds for every *bounded* $\mathfrak{B}$-measurable numerical function, then T is called *boundedly complete* for ω. The following implications are true [8].

$$(2) \qquad \text{Completeness} \Rightarrow \text{bounded completeness} \Rightarrow \text{minimality.}$$

[3] The term *minimal* was introduced by Lehmann and Scheffé [8], whereas Bahadur [2] describes the same concept with the term *necessary and sufficient statistic*

Suppose a composite hypothesis H specifies $\theta \varepsilon \omega \subset \Omega$. We shall consider randomized tests for H with test functions ϕ, where, for each $x \varepsilon \mathfrak{X}$, $0 \leqq \phi(x) \leqq 1$, ϕ measurable, and H is rejected with probability $\phi(x)$ if x is observed. Among all tests we restrict ourselves to similar tests, defined by the condition that $E_\theta \phi$ is independent of θ if $\theta \varepsilon \omega$. If T is a sufficient statistic for ω, $\mathfrak{A}_0 \subset \mathfrak{A}$ its sufficient sub σ-field and ϕ any test, we can consider the $\mathfrak{A}_0$-measurable function $E(\phi \mid \mathfrak{A}_0)$. If α is a number, $0 < \alpha < 1$, and if ϕ is such that $E(\phi \mid \mathfrak{A}_0) \equiv \alpha$, then clearly $E_\theta \phi = \alpha$ for all $\theta \varepsilon \omega$, so that ϕ is similar. Such a ϕ is called a *test of Neyman structure* [8]. If T is a boundedly complete sufficient statistic for ω, then every similar test has Neyman structure [8]. On the other hand, if a sufficient statistic T is not boundedly complete for ω, then there exist similar tests which do not have Neyman structure. This follows from the fact that the bounded incompleteness implies the existence of a $\mathfrak{B}$-measurable numerical function g on $\mathfrak{I}$, bounded below by $-\alpha$, above by $1 - \alpha$, different from 0 on a set of positive probability (with respect to $\mathcal{P}_\omega$), with $E_\theta g(T) = 0$ for all $\theta \varepsilon \omega$. With f on $\mathfrak{X}$ defined by $f(x) = g(T(x))$, we have that $\phi = f + \alpha$ is similar of size α, but $E(\phi \mid \mathfrak{A}_0) - \alpha = f \neq 0$ on a set of positive probability, so ϕ is not a test of Neyman structure. Conversely, for any similar test ϕ we can form the function $f = E(\phi \mid \mathfrak{A}_0) - \alpha$ and define g on $\mathfrak{I}$ by $g(T(x)) = f(x)$, so that

$$E_\theta g(T) = 0$$

for all $\theta \varepsilon \omega$. It follows that all similar tests can be found by constructing all bounded numerical functions g on $\mathfrak{I}$ whose expectations vanish for all $\theta \varepsilon \omega$.

3. The D method for constructing similar tests in the case of a family of regular exponential densities. In this section the restriction of θ to ω will be understood. Let the distribution of T, induced by P_θ, have a density with respect to m-dimensional Lebesgue measure, and let this density p_θ be of the form

$$(3) \qquad p_\theta(t) = c(\theta) \exp\left[-\sum_{i=1}^{m} s_i(\theta) t_i\right] h(t)$$

in which $t = (t_1, \cdots, t_m)$, and $s_1, \cdots, s_m$ are real valued functions on ω. We shall assume that the function h is of such a nature that it is possible to find a closed m-dimensional cube C on which h is bounded away from 0. With this restriction on h, the family (3) will be called a family of *regular exponential densities*. Exponential densities which arise in statistics are always regular.

If ω is an m-dimensional subset of an m-dimensional Euclidean space, then, under mild conditions, T with density (3) is complete for ω [9]. In that case every similar test has Neyman structure. From the point of view of the present paper the interesting case arises when ω is a subset of an $m - 1$ dimensional Euclidean space. In that case θ has at most $m - 1$ components, so that the m functions s_i are functions of at most $m - 1$ parameters. Eliminating those parameters will result in a functional relation between the s_i. Suppose that this relation can be put in the form

$$(4) \qquad P(s_1, \cdots, s_m) = 0$$

in which P is a polynomial of positive degree in at least one of the s_i. It should be kept in mind that (4) holds identically in θ.

As discussed in section 2, a similar test of non-Neyman structure can be constructed by constructing a bounded function g on $\mathfrak{I}$, $g \neq 0$ on a set of positive probability, such that

$$\int g(t)p_\theta(t)\, dt \underset{\theta}{\equiv} 0 \tag{5}$$

Using (3), remembering that h is bounded away from 0 on some m-dimensional cube C, it suffices to construct a bounded function F which is $\neq 0$ on a subset of C of positive Lebesgue measure, vanishes outside C, and satisfies

$$\int F(t) \exp\left[-\sum_1^m s_i(\theta)t_i\right] dt \underset{\theta}{\equiv} 0 \tag{6}$$

The function g in (5) can then be taken as F/h. The left hand side of (6) is the m-dimensional Laplace transform of F, denoted by $\mathfrak{L}(F)$·

$$\int F(t_1 \cdots, t_m) \exp\left[-\sum_1^m s_i t_i\right] dt = \mathfrak{L}(F)(s_1 \cdots, s_m). \tag{7}$$

The problem is to construct F in such a way that $\mathfrak{L}(F) = 0$ for all values of $s(\theta)$, $\theta \in \omega$. This can be done with help of (4). Let P be of degree d and let G be a function on $\mathfrak{I}$ possessing all partial derivatives of dth order in the interior of C, vanishing outside C, and having all partial derivatives of $d - 1$st order continuous on the boundary of C. An example of such a function is the following. Let C be given by $a_i \leqq t_i \leqq a_i + l$ $(i = 1, \cdots, m)$, then on C we can take $G(t) = \prod_{i=1}^m (t_i - a_i)^d(a_i + l - t_i)^d$. Now denote by D the differential operator

$$D = P\left(\frac{\partial}{\partial t_1}, \cdots, \frac{\partial}{\partial t_m}\right). \tag{8}$$

We then have

$$\mathfrak{L}(DG)(s) = P(s)\mathfrak{L}(G)(s) \tag{9}$$

in which $s = (s_1, \cdots, s_m)$. Since the right hand side of (9) is $\equiv 0$ by (4), we may take F in (7) to be $F = DG$. The final result is therefore

$$g(t) = (DG(t))/h(t) \tag{10}$$

for suitably chosen G, and

$$\phi(t) = \alpha + (DG(t))/h(t) \tag{11}$$

is a size α similar test of non-Neyman structure.

Even for one m-dimensional cube C the number of choices for G is large. In addition there will usually be a large number of m-dimensional cubes on each of which h is bounded away from 0, and finally one may consider regions other than cubes for which the construction of functions G is possible. Thus, there will be a large class of functions g satisfying (5) which can be generated by the

differential operator method, called the *D method* henceforth. Whether this method, in general, will give all those functions g, is still an open question. In one particular case the question has been answered in the affirmative, provided the definition of D method is taken sufficiently wide (see section 6).

Suppose that with the help of the D method a similar test $\phi(T)$ is constructed, and that it is desired to consider similar tests which do not necessarily depend on T only. Let ψ be a test function defined on the sample space $\mathfrak{X}$. If ψ is chosen to satisfy $E(\psi \mid t) = \phi(t)$, then ψ is also similar. In particular, it will usually be possible to construct in this way a similar rejection region w, in which case ψ is the indicator of w (this construction fails is $\mathfrak{X}$ is a subspace of a Euclidean space with same dimension as $\mathfrak{J}$) A similar region w is constructed by demanding

$$P(w \mid t) = \phi(t). \tag{12}$$

In other words, on each surface $T = t$ in the sample space a region is selected which has conditional probability $\phi(t)$. This generalizes the construction of a similar region of Neyman structure [10]. Equation (12) will be used in section 4, example 2.

4. Examples of the D method. EXAMPLE 1 (Behrens-Fisher problem). Let $X_1, \cdots, X_{n_1}$ be n_1 independent observations on a normal variable with mean μ_1, variance σ_1^2, and $Y_1, \cdots, Y_{n_2}$, n_2 independent observations on a normal variable with mean μ_2, variance σ_2^2. The X's and Y's are independent, and all parameters are unknown. Under the hypothesis tested, which is $\mu_1 = \mu_2$, the joint distribution of the X's and Y's has an exponential density with exponential factor

$$\exp\left[-\frac{1}{2\sigma_1^2}\sum_1^{n_1} x_i^2 + \frac{\mu}{\sigma_1^2}\sum_1^{n_1} x_i - \frac{1}{2\sigma_2^2}\sum_1^{n_2} y_j^2 + \frac{\mu}{\sigma_2^2}\sum_1^{n_2} y_j\right]$$

in which μ is the common value of μ_1 and μ_2. We may take

$$T_1(x) = \sum_1^{n_1} x_i^2, \qquad T_2(x) = \sum_1^{n_1} x_i, \qquad T_3(x) = \sum_1^{n_2} y_j^2, \qquad T_4(x) = \sum_1^{n_2} y_j,$$

$$s_1(\theta) = \frac{1}{2\sigma_1^2}, \qquad s_2(\theta) = \frac{-\mu}{\sigma_1^2}, \qquad s_3(\theta) = \frac{1}{2\sigma_2^2}, \qquad s_4(\theta) = \frac{-\mu}{\sigma_2^2}.$$

The s_i are linearly independent, from which it can be shown that

$$T = (T_1, T_2, T_3, T_4)$$

is a minimal sufficient statistic for ω. T has a regular exponential density of form (3), with

$$h(t) = (n_1t_1 - t_2^2)^{(n_1-3)/2}(n_2t_3 - t_4^2)^{(n_2-3)/2} \tag{13}$$

if $n_1t_1 \geqq t_2^2$, $n_2t_3 \geqq t_4^2$, and $h(t) = 0$ otherwise. By eliminating μ, σ_1, σ_2 from

the four s_i we obtain $s_1 s_4 - s_2 s_3 = 0$ as a realization of (4). The differential operator D in (8) is then

(14) $$D = \frac{\partial}{\partial t_1}\frac{\partial}{\partial t_4} - \frac{\partial}{\partial t_2}\frac{\partial}{\partial t_3}$$

and for suitably chosen G the test

(15) $$\phi(t) = \alpha + h^{-1}(t)\left(\frac{\partial}{\partial t_1}\frac{\partial}{\partial t_4} - \frac{\partial}{\partial t_2}\frac{\partial}{\partial t_3}\right)G(t)$$

is similar and of size α, where $h(t)$ is given by (13) Whether this method can be used to show the existence of an invariant similar region, such as the one proposed by Welch [1], [15], has not yet been investigated.

It should perhaps be mentioned here that the approach to the Behrens-Fisher problem by Wald [14] is essentially different, since Wald does not require the test to be similar.

Example 2. (Standardized mean of a normal population). Suppose we make $n + 1$ independent observations on a normal variable and consider hypotheses concerning the ratio of mean to standard deviation By an orthogonal transformation this problem can be brought in the following form: Let $X_0, \cdots, X_n$ be independent and normal, with common, unknown variance σ^2. X_0 has unknown mean μ, $X_1, \cdots, X_n$ have mean 0 Denote $\mu/\sigma = r$, then for some given r_0 the hypothesis tested is $r = r_0$. For the time being the alternative to be considered is immaterial. For later reference, however, suppose that the alternative is $r > r_0$. We then have

$$\Omega = \{(r, \sigma): r \geqq r_0, \sigma > 0\},$$

$$\omega = \{(r, \sigma): r = r_0, \sigma > 0\}.$$

Under the hypothesis the joint distribution of the X_i has the form given by (3), with exponential factor

$$\exp\left[-\frac{1}{2\sigma^2}\sum_0^n x_i^2 + \frac{r_0}{\sigma}x_0\right]$$

so that we may take

$$T_1(x) = \sum_0^n x_i^2, \quad T_2(x) = x_0, \quad s_1(\sigma) = \frac{1}{2\sigma^2}, \quad s_2(\sigma) = -\frac{r_0}{\sigma}.$$

$T = (T_1, T_2)$ is minimal sufficient, since s_1 and s_2 are linearly independent. Elimination of σ from s_1 and s_2 gives $s_2^2 - 2r_0^2 s_1 = 0$, so that we can take

(16) $$P(s_1, s_2) = s_2^2 - 2r_0^2 s_1$$

and

(17) $$D = \frac{\partial^2}{\partial t_2^2} - 2r_0^2\frac{\partial}{\partial t_1}$$

The function h in (3) is found to be

$$h(t_1, t_2) = (t_1 - t_2^2)^{(n/2)-1} \tag{18}$$

if $t_1 \geqq t_2^2$, and $h = 0$ otherwise. For suitably chosen $G(t_1, t_2)$ the test function

$$\phi(t_1, t_2) = \alpha + (t_1 - t_2^2)^{1-n/2}\left(\frac{\partial^2}{\partial t_2^2} - 2r_0^2\frac{\partial}{\partial t_1}\right)G(t_1, t_2) \tag{19}$$

is similar and of size α.

Equation (19) can be used to demonstrate the existence of similar tests which are not invariant. In the present problem an invariant test is a function of $T_2/\sqrt{T_1}$ only. Choose for G in (19) the following function:

$$G(t_1, t_2) = c(t_1 - t_2^2)^{(n/2)+1}e^{-t_1} \tag{20}$$

if $t_1 \geqq t_2^2$ and $G = 0$ otherwise, with $c > 0$ chosen so small that ϕ is bounded between 0 and 1. It is easily checked that after substitution into (19) the resulting test function is not invariant. This example can also be used to show the existence of similar rejection regions which are not equivalent to a cone in the sample space (we shall call two tests *equivalent* if they have the same power function, and by a *cone* is meant a union of rays through the origin). If w is any rejection region, ϕ the corresponding test function, given by (12), then w and ϕ are equivalent since T is sufficient, not only for ω, but also for Ω. If w_1 is a cone in the sample space, then the corresponding ϕ_1 is invariant. Let w_2 be any rejection region equivalent to w_1, ϕ_2 the corresponding test function; then ϕ_1 and ϕ_2 are equivalent. Now T is not only sufficient for Ω, it is also complete for Ω. Since ϕ_1 and ϕ_2 have the same power functions, it follows then that $\phi_1 = \phi_2$ a.e. and thus ϕ_2 is also invariant. The existence of a noninvariant similar test ϕ implies then the existence of a similar region which is not equivalent to any cone in the sample space.[4]

5. A criterion for bounded incompleteness in the case of regular exponential densities. Let the family of distributions be given by (3), with $\theta \in \omega$. By (2), if T is not minimal sufficient for ω, then T cannot be boundedly complete. This happens, for instance, if the s_i are linearly dependent on ω because the exponent $-\Sigma s_i t_i$ in (3) can then be written as a linear combination of fewer than m of the t_i. The incompleteness in this case also follows from the applicability of the D method of section 3, because of the existence of a polynomial P, linear in this case, for which (4) holds. On the other hand, if the m functions s_i are linearly independent on ω, then T is minimal sufficient for ω. Even if this is the case, T may still be boundedly incomplete. Theorem 2 below tells when this will happen. Its proof uses the D method of section 3. The conditions of Theorem 2 are designed to guarantee the existence of the polynomial P on the left

[4] This seems to contradict a statement by Patnaik [12] to the effect that in the problem under consideration every similar region is equivalent to a cone in the sample space. However, Patnaik's proof is unconvincing, and the non-invariant ϕ exhibited above provides a counter example.

hand side of (4), such that P is not the zero polynomial. This is made possible by the following theorem, due to A. Seidenberg (private communication). The proof is given in Appendix 1.

THEOREM 1. (Seidenberg). *Let for each* i, $i = 1, \cdots, m$, $P_i(s_i; \theta_1, \cdots, \theta_k)$ *be a polynomial in* s_i *and the* $\theta_j (j = 1, \cdots, k)$, *with coefficients in some field* K, *where* $k < m$ *and* P_i *is of positive degree in* s_i. *Let* $A_i(\theta)$ *be the leading coefficient of* P_i *as a polynomial in* s_i. *Then there is a polynomial* $P(s_1, \cdots, s_m)$ *with coefficients in* K, *which is not the zero polynomial, and a power product* $B(\theta)$ *of the* $A_i(\theta)$, *such that* $B(\theta)P(s) = 0$ *whenever* $P_i = 0$ *for all* i.

COROLLARY. *If* θ *is restricted to a set* Θ, *and if, for each* $\theta \varepsilon \Theta$ *and each* i,

$$A_i(\theta) \neq 0,$$

then $P = 0$ *whenever* $P_i = 0$ *for all* i.

For, if $A_i(\theta) \neq 0$, $i = 1, \cdots, m$, then $B(\theta) \neq 0$.

In the application we want to make of the corollary, the set Θ is ω. Furthermore, we shall assume the s_i of section 3 to be algebraic functions of the θ_j, for $\theta \varepsilon \omega$. Then for each i there is a polynomial P_i in s_i and the θ_j such that $P_i(s_i; \theta_1, \cdots, \theta_k) = 0$ if $\theta \varepsilon \omega$. We shall further assume that the $A_i(\theta)$ are $\neq 0$ if $\theta \varepsilon \omega$. These conditions will be satisfied in particular if, for each i, s_i on ω is a rational function of the θ_j, with nonvanishing denominator.

THEOREM 2. *Suppose a family of regular exponential densities is given by* (3), *with* $\theta \varepsilon \omega$; ω *is a subset of a* k*-dimensional Euclidean space, with* $k < m$; *on* ω, *the* m *functions* s_i *are algebraic functions of the* k *parameters* θ_j, *so that*

$$P_i(s_i; \theta_1, \cdots, \theta_k) = 0$$

for some polynomial P_i $(i = 1, \cdots, m)$; $A_i(\theta)$, *the leading coefficient of* P_i *as a polynomial in* s_i, *does not vanish anywhere on* ω *for any* i. *Then* T *is boundedly incomplete for* ω.

The proof follows immediately from the constructibility, by the D method of section 3, of a bounded function g, $g \neq 0$ on a set of positive probability, satisfying $E_\theta g(T) = 0$ for all $\theta \varepsilon \omega$.

In both examples in section 4 the s_i are rational functions of the θ_j, with nonvanishing denominators, and in both cases $k = m - 1 < m$, so that Theorem 2 applies. This provides another proof of the well-known fact that in the Behrens-Fisher problem, as well as in the problem of testing the ratio of mean to standard deviation in a normal population, the minimal sufficient statistic is boundedly incomplete.

It would be interesting to know how much the assumptions of Theorem 2 can be relaxed. It is certainly not necessary that the s_i be algebraic functions of the θ_j, for, if $m = 2$, $k = 1$, $s_1 = \cos\theta$, $s_2 = \sin\theta$, then $s_1^2 + s_2^2 - 1 = 0$, as a realization of (4), so that the D method applies. It is not even necessary for incompleteness that there exists a polynomial P in the s_i which vanishes for all $\theta \varepsilon \omega$, as the next example will show. Take $m = 2$, $k = 1$, $s_1 = -\ln\theta$, $s_2 = -\ln(1 - \theta)$, with $0 < \theta < 1$. Instead of (4) we have a transcendental equation:

$\exp[-s_1] + \exp[-s_2] - 1 = 0$. With help of this equation one can easily construct functions F of the kind mentioned in section 3. For example, the function F whose 2-dimensional Laplace transform is

$$\mathcal{L}(F)(s_1, s_2) = \frac{1}{s_1 s_2}(e^{-s_1} + e^{-s_2} - 1)(e^{-a_1 s_1} - e^{-b_1 s_1})(e^{-a_2 s_2} - e^{-b_2 s_2})$$

is bounded between -1 and 1, vanishes outside the rectangle

$$a_1 \leqq t_1 \leqq b_1 + 1, \qquad a_2 \leqq t_2 \leqq b_2 + 1,$$

and has vanishing Laplace transform for all θ between 0 and 1. On the other hand, the fact that $k < m$ is not sufficient for bounded incompleteness, nor is the additional restriction of analyticity of the s_i sufficient. The following example is due to L. J. Savage (private communication). In (3) choose $m = 2$, $k = 1$, $s_1 = \theta \cos \theta$, $s_2 = \theta \sin \theta (\theta > 0)$, $h(t) = 1$ for t in some square, $h = 0$ otherwise. Here the s_i are analytic functions of θ, but yet it can be shown that the family of distributions is complete. Another example is due to D. L. Burkholder (private communication) and differs from Savage's example only in that $s_1 = \theta \cos (1/\theta)$, $s_2 = \theta \sin (1/\theta)$. This example is a little less regular than Savage's example, but on the other hand the completeness of the family of distributions is easier to show.

6. Completeness of the D method in the case of a hypothesis concerning the standardized mean of a normal population. In this section it will be shown that in Example 2 of section 4 all similar tests can be generated by the D method, provided the D method is defined in a sufficiently broad manner. That is, we want to show that for each similar test ϕ there exists a function G satisfying (19) and certain other conditions. In section 3 the functions G were restricted to some m-dimensional cube on which h is bounded away from 0 but it was remarked there that this restriction is not necessary. We shall not even demand that $G = 0$ whenever $h = 0$. In fact, the main thing of importance was the validity of (9), and even this we shall relax slightly in the problem under consideration.

Equation (19) can be put in the form

$$\left(\frac{\partial}{\partial t_1} - \frac{1}{2r_0^2}\frac{\partial^2}{\partial t_2^2}\right) G = \frac{\sqrt{2\pi}}{r_0}\varphi \tag{21}$$

where φ is defined by

$$\varphi(t) = -(\sqrt{8\pi}r_0)^{-1}h(t)(\phi(t) - \alpha). \tag{22}$$

Equation (21) can be considered as the heat equation in one dimension, if t_1 is interpreted as time, t_2 as position, G as temperature, and $(\sqrt{2\pi}/r_0)\varphi$ as a heat source, capable of producing both positive and negative heat, whose strength and spatial distribution varies with time. If this were an actual heat problem, its solution could be written down at once, employing the usual Green's function for the heat operator:

(23) $$G(t_1, t_2) = \int\int \varphi(t_1', t_2')(t_1 - t_1')^{-1/2} \exp\left[-\frac{r_0^2}{2}\frac{(t_2 - t_2')^2}{t_1 - t_1'}\right] dt_1'\, dt_2'$$

where the integration is over the strip $0 \leqq t_1' \leqq t_1$. Since $h(t')$, and therefore, $\varphi(t')$, is zero unless $t_2'^2 \leqq t_1'$, we may integrate over $t_2'^2 \leqq t_1' \leqq t_1$. The question to be answered next is whether, and if so, in what sense, the formal solution (23) to (21), and therefore, to (19) is a representation of ϕ.

We shall at once study the power function of any similar test ϕ, since some of the results are needed in section 7. Let Ω and ω be as defined in section 1, Example 2. We shall assume $r_0 > 0$. As remarked in section 4, the statistic

$$T = (T_1, T_2)$$

is sufficient for Ω, and it suffices therefore to consider test functions ϕ which depend only on T. The power function of ϕ is $\beta(r, \sigma) = E_{r,\sigma}\phi(T_1, T_2)$. Suppose ϕ satisfies (19), then we get after substitution:

(24) $$\beta(r, \sigma) = \alpha + c(r, \sigma) \int\int \cdot \exp\left[-\frac{1}{2\sigma^2}t_1 + \frac{r}{\sigma}t_2\right]\left(\frac{\partial^2}{\partial t_2^2} - 2r_0^2\frac{\partial}{\partial t_1}\right) G(t_1, t_2)\, dt_1\, dt_2$$

where the integration is over $0 \leqq t_1 < \infty$, $-\infty < t_2 < \infty$. We may effect this integration by taking the upper limits on t_1 and t_2 as A, B respectively, and then let $A \to \infty$, $B \to \infty$ in any order. With respect to the types of functions G to be considered it will not be necessary to do something similar with the lower limit on t_2. If the upper limits on t_1 and t_2 are A and B, one can integrate by parts, obtaining an integral

(25) $$\frac{r^2 - r_0^2}{\sigma^2}\int_0^A dt_1 \int_{-\infty}^B G(t_1, t_2) \exp\left[-\frac{1}{2\sigma^2}t_1 + \frac{r}{\sigma}t_2\right] dt_2$$

plus the following integrated terms:

(26) $$-2r_0^2 \int_{-\infty}^B G(A, t_2) \exp\left[-\frac{1}{2\sigma^2}A + \frac{r}{\sigma}t_2\right] dt_2$$

(27) $$-\frac{r}{\sigma}\int_0^A G(t_1, B) \exp\left[-\frac{1}{2\sigma^2}t_1 + \frac{r}{\sigma}B\right] dt_1$$

(28) $$\int_0^A \frac{\partial G(t_1, B)}{\partial t_2} \exp\left[-\frac{1}{2\sigma^2}t_1 + \frac{r}{\sigma}B\right] dt_1$$

There is also an integral involving G on the t_2-axis. For any G given by (23), $G(0, t_2) = 0$, so that the integral mentioned in the preceding sentence vanishes trivially. It is sufficient, then, to consider only functions G which vanish if $t_1 = 0$. Now if G is given by (23), with φ defined by (22) and ϕ similar of size α, then it can be shown that (26)-(28) vanish in the limit if we let first $B \to \infty$ and then $A \to \infty$. A proof is given in Appendix 2. Using (24) and (25) it follows that

$$\beta(r, \sigma) = \alpha + c(r, \sigma) \frac{r^2 - r_0^2}{\sigma^2} \lim_{A\to\infty} \lim_{B\to\infty} \int_0^A dt_1 \int_{-\infty}^B G(t_1, t_2) \cdot \exp\left[-\frac{1}{2\sigma^2} t_1 + \frac{r}{\sigma} t_2\right] dt_2 \tag{29}$$

We see from (29) that $\beta(r_0, \sigma) = \alpha$ identically in σ, as it should.

The reason we could get the power function in the form (29) is that in this problem the density of T is of the exponential form (3) on the whole of Ω. The exponent of the exponential factor is $-t_1/(2\sigma^2) + rt_2/\sigma$, so that on Ω we have $s_1 = 1/(2\sigma^2)$, $s_2 = -r/\sigma$. The polynomial (16) is now defined on the whole of Ω:

$$P(s) = s_2^2 - 2r_0^2 s_1 = \frac{r^2 - r_0^2}{\sigma^2} \tag{30}$$

(On ω, $r = r_0$, so $P = 0$ as it should). We made the integrated terms (26)–(28) vanish by taking limits in a special way. This suggests, for this problem to redefine the 2-dimensional Laplace transform as follows:

$$\mathfrak{L}(F)(s_1, s_2) = \lim_{A\to\infty} \lim_{B\to\infty} \int_0^A dt_1 \int_{-\infty}^B F(t_1, t_2) \exp[-s_1 t_1 - s_2 t_2]\, dt_2 \tag{31}$$

With P and $\mathfrak{L}$ defined by (30) and (31), we have proved that if ϕ is similar, and G is the corresponding function given by (23), then (9) is valid on the whole of Ω. Adding α to both sides of (9) then produces (29).

In order to characterize the whole class of similar tests, consider the class $\mathfrak{C}$ of functions G defined on the right half (t_1, t_2) plane which satisfy the following conditions (with D defined by (17)):

(i) $DG(t_1, t_2) = 0$ if $t_2^2 > t_1$,

(ii) $-\alpha \leqq (t_1 - t_2^2)^{1-n/2} DG(t_1, t_2) \leqq 1 - \alpha$ if $t_2^2 \leqq t_1$,

(iii) $G = 0$ if $t_1 = 0$, and $G(t_1, t_2) \to 0$ as $t_2 \to -\infty$, for each t_1,

(iv) The integrals (26)–(28) approach 0 if we let first $B \to \infty$ and then $A \to \infty$.

For every similar size α test function ϕ there is, by (23) and (22), a unique G, satisfying the conditions (i)–(iv), so that $G \varepsilon \mathfrak{C}$. Conversely, for any $G \varepsilon \mathfrak{C}$ we have shown that ϕ given by (19) is similar and of size α. Thus, there is a one-to-one correspondence between the members of $\mathfrak{C}$ and the similar size α test functions. The class $\mathfrak{C}$ gives therefore a complete characterization of the similar tests. Unfortunately, condition (iv) is not a very easy one. There is an important subclass of $\mathfrak{C}$ where (iv) is obviously fulfilled, consisting of those functions G in $\mathfrak{C}$ which vanish identically for $t_2 > \sqrt{t_1}$. This is the case, for instance, with all functions G leading to an invariant test. For a proof of this fact see Appendix 3. It would be desirable if (iv) could be replaced by a simpler condition. The possibility is not excluded that conditions (i)–(iv) imply that $G(t_1, t_2) = 0$ for all $t_2 > \sqrt{t_1}$, but whether this is so is an open question.

7. Some remarks on the search for an optimum test in the problem of section 6. Consider the class $\mathcal{C}$ defined in section 6. Let ϕ_1, ϕ_2 be two similar size α tests, G_1, G_2 the corresponding functions in $\mathcal{C}$, and β_1, β_2 their power functions. It follows from (29), since $r^2 \geqq r_0^2$, that if $G_1 \geqq G_2$, then $\beta_1 \gtreqless \beta_2$, so that ϕ_1 is uniformly more powerful than ϕ_2. Since every similar ϕ has a representative $G \varepsilon \mathcal{C}$, if there would exist a $G_0 \varepsilon \mathcal{C}$ such that $G_0 \geqq G$ *for every* $G \varepsilon \mathcal{C}$, *then the* test function ϕ_0 corresponding to G_0 would be UMP (uniformly most powerful) among all similar tests. To decide whether or not such a dominating function G_0 exists, the following observations may be of help. The first observation is that in the problem under consideration every invariant test—that is, depending only on $T_2/\sqrt{T_1}$—is similar. Secondly, if we denote by $\mathcal{C}^*$ the subclass of $\mathcal{C}$ representing invariant tests, then in $\mathcal{C}^*$ there is a function G_0^* which dominates every $G^* \varepsilon \mathcal{C}^*$. The corresponding test function ϕ_0^* is therefore UMP among all invariant tests. ϕ_0^* is nonrandomized, with a rejection region of the form $t_2/\sqrt{t_1} >$ constant That ϕ_0^* is UMP invariant is a known result [12], obtainable more directly by the observation that $T_2/\sqrt{T_1 - T_2^2}$ has a noncentral t-distribution with a monotonic likelihood ratio [3], [4], [7] The third observation we want to make is that if the dominating function G_0 exists, it has to coincide with G_0^*. This follows from the following proposition. *If a* UMP *similar test based on* T *exists, it is necessarily invariant* The analogous statement, with "similar" replaced by "unbiased," is well known [5], [6]. In fact, both statements are special cases of the following more general theorem, due to E. L. Lehmann (private communication)· *Let* $\mathcal{G}$ *be a group of transformations which leaves the problem invariant, and let* $\mathcal{K}$ *be a class of tests which is closed under* $\mathcal{G}$ *If there is a unique* UMP *test in* $\mathcal{K}$, *it is almost invariant* (The uniqueness is understood to be a.e.). The proof of this theorem follows the same lines as in the special case that $\mathcal{K}$ is the class of unbiased tests of fixed size In our problem $\mathcal{K}$ is the class of similar tests of size α, based on T. $\mathcal{K}$ is clearly closed under $\mathcal{G}$. If there is a UMP test in $\mathcal{K}$, its uniqueness follows from the completeness of T for Ω. Finally, in our problem an almost invariant function can be shown to be invariant (see also [17], footnote 3, and [5]).

The conclusion drawn from the preceding discussion is that there is a dominating function $G_0 \varepsilon \mathcal{C}$ if and only if G_0^* is the dominating function. Whether or not this is so is still an open question, and consequently, *it is still unknown* whether a UMP similar test exists. A last remark may be added to this. As remarked in section 6 and proved in Appendix 3, the functions G^* in $\mathcal{C}^*$ have the remarkable property that they vanish for $t_2 \geqq \sqrt{t_1}$. This property holds then in particular for G_0^*. Taking into account that $G \varepsilon \mathcal{C} \Rightarrow -aG \varepsilon \mathcal{C}$ for sufficiently small $a > 0$, we conclude that if G_0^* is a dominating function in $\mathcal{C}$, then every $G \varepsilon \mathcal{C}$ must also have the property $G(t_1, t_2) = 0$ if $t_2 \geqq \sqrt{t_1}$. If this were indeed true, then condition (iv) in section 6 could be replaced by the much simpler condition $G(t_1, t_2) = 0$ if $t_2 \geqq \sqrt{t_1}$. However, as remarked in section 6, even this property has not yet been proved.

Acknowledgements. The writer is highly indebted to Dr. Erich L. Lehmann, whose inspiring lectures laid the foundation for this study, and whose constructive criticism was always deeply appreciated. A special word of thanks is due Dr. Abraham Seidenberg for making available Theorem 1 in section 5. Gratitude also goes to Dr. Jerzy Neyman for continuous interest in this work. Finally, the writer wants to thank Dr. J. Kiefer and the referee for many valuable suggestions.

REFERENCES

[1] ALICE A. ASPIN, "An examination and further development of a formula arising in the problem of comparing two mean values," *Biometrika*, Vol. 35 (1948), pp. 88-96.
[2] R. R. BAHADUR, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 423–462.
[3] S. KARLIN, "Decision theory for Pólya type distributions. Case of two actions, I." *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*," Vol. 1, University of California Press, 1956, pp. 115–128.
[4] W. KRUSKAL, "The monotonicity of the ratio of two non-central t density functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 162–165.
[5] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1–26.
[6] E. L. LEHMANN, "Theory of testing hypotheses," mimeographed notes recorded by Colin Blyth, ASUC Bookstore, University of California, Berkeley, 1949.
[7] E. L. LEHMANN, "Ordered families of distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 399–419.
[8] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions, and unbiased estimation. Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.
[9] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions, and unbiased estimation. Part II," *Sankhyā*, Vol. 15 (1955), pp. 219–236.
[10] J. NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc., London, Series A*, Vol. 236 (1937), pp. 333–380.
[11] J. NEYMAN AND E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc., London, Series A*, Vol. 231 (1933), pp. 289–337.
[12] P. B. PATNAIK, "Hypotheses concerning the means of observations in normal samples," *Sankhyā*, Vol. 15 (1955), pp. 343-372.
[13] H. SCHEFFÉ, "On solutions of the Behrens-Fisher problem, based on the t-distribution," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 35–44.
[14] A. WALD, "Testing the difference between the means of two normal populations with unknown standard deviations," *Selected Papers in Statistics and Probability*, Stanford University Press, 1957, pp. 669-695.
[15] B. L. WELCH, "On the studentization of several variances," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 118-122.
[16] R. A. WIJSMAN, "Incomplete sufficient statistics and similar tests," (abstract), *Ann. Math. Stat.*, Vol. 27 (1956), p. 864.
[17] J. WOLFOWITZ, "The power of the classical tests associated with the normal distribution," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 540–551.

Appendix 1. PROOF OF THEOREM 1 (Seidenberg). For the purpose of this proof we shall replace the s_i by x_i. Let $P_i = A_i(\theta)x_i^{d_i} + \cdots$. Let $d = \max\{d_i\}$. Multiplying P_i by $x_i^{d-d_i}$, we may suppose all the d_i equal. Multiplying P_1 by

$A_2 \cdot \cdots \cdot A_m$, P_2 by $A_1 A_3 \cdot \cdots \cdot A_m$, etc., we may suppose all the A_i equal. Now we have $P_i = \Delta(\theta)x_i^d + \cdots$, $i = 1, \cdots, m$, where Δ is some polynomial in $\theta_1, \cdots, \theta_k$.

Suppose we have a congruence of the form

$$\Pi x_i^{r_i} \Delta^{\Sigma r_i} \equiv R(x, \theta) \bmod (P_1, \cdots, P_m)$$

(i.e. the two sides are equal whenever all P_i vanish), with R a polynomial in the x's and θ's. Let $M = \max \{\deg_\theta P_i\}$. The left hand side has degree in the θ's at most $M\Sigma r_i$. Assume this to be the case also for $R(x, \theta)$. Assume further that $\deg_{x_i} R(x, \theta) \leqq d - 1$, $i = 1, \cdots, m$ Multiplying the congruence by $x_j\Delta$, on the left we get a power product of degree $1 + \Sigma r_i$ in the x_i times $\Delta^{1+\Sigma r_i}$. On the right there possibly appears a power x_j^d: if so, we replace Δx_j^d by

$$(\Delta x_j^d - P_j) \bmod P_j$$

In this way we get a congruence

$$\Pi x_i^{s_i} \Delta^{\Sigma s_i} \equiv R'(x, \theta) \bmod (P_1, \cdots, P_m)$$

with $\Sigma s_i = 1 + \Sigma r_i$, $\deg_{x_i} R' \leqq d - 1$ $(i = 1, \cdots, m)$, $\deg_\theta R' \leqq M\Sigma s_i$. The congruences

$$x_i^d \Delta^d \equiv \Delta^{d-1}(x_i^d - P_i) \bmod (P_1, \cdots, P_m)$$

are of the above form. Multiplying by various power products of the $x_j\Delta$, we again get congruences of the stated form. Let $s \geqq s_0 = m(d - 1) + 1$. Then any power product of the x_i of degree s must have a factor x_i^d for at least one i. Hence we can get a congruence of the desired form with any power product of the $x_j\Delta$ of degree s on the left. For any such power product there may be several congruences: choose one.

For a fixed integer $\gamma \geqq s_0$ (to be determined in a moment), we consider all the power products of the $x_i\Delta$ of degree between s_0 and γ; and all the congruences, one for each power product. We still multiply each of these by an appropriate power of Δ so that Δ^γ is the power of Δ occurring on the left. On the right, then, all polynomials are of degree $\leqq M\gamma$ in the θ's and of degree $\leqq d - 1$ in each x_i.

Let $N(p, q)$ be the number of distinct power products of degree p or less in q letters. Then $N(p, q) = (p + q)(p + q - 1) \cdots (p + 1) / q!$ We are considering, then, $N(\gamma, m) - N(s_0 - 1, m)$ congruences. The right hand sides of these congruences are linear combinations over K of power products of degree $\leqq M\gamma$ in the θ's and of degree $\leqq m(d - 1)$ in the x's; therefore in at most $N(M\gamma, k)N(m(d - 1), m)$ power products. Since

$$\deg_\gamma [N(\gamma, m) - N(s_0 - 1, m)] = m > k = \deg_\gamma N(M\gamma, k)N(m(d - 1), m),$$

we see that for sufficiently large γ,

$$N(\gamma, m) - N(s_0 - 1, m) > N(M\gamma, k)N(m(d - 1), m).$$

Let γ be taken large enough for this to be realized. Then there exist $c_{i_1,\cdots,i_m} \varepsilon K$, not all $= 0$, such that

$$\Delta^\gamma \Sigma c_{i_1,\cdots,i_m} x_1^{i_1} \cdot \cdots \cdot x_m^{i_m} \equiv 0 \bmod (P_1, \cdots, P_m). \qquad \text{Q.E.D.}$$

Appendix 2. It will be proved here that the integrated terms (26)–(28) vanish in the limit $B \to \infty$, then $A \to \infty$. Since ϕ is similar and of size α, the function φ defined by (22) has the property

$$\int_0^\infty \int_{-\sqrt{t_1}}^{\sqrt{t_1}} \varphi(t_1, t_2) \exp\left[-\frac{1}{2\sigma^2} t_1 + \frac{r_0}{\sigma} t_2\right] dt_2\, dt_1 \underset{\sigma}{\equiv} 0. \tag{32}$$

This property is crucial for showing (26) $\to 0$, but is not needed for (27) and (28).

We shall first treat (26). Since σ is an arbitrary positive number, we shall give the proof with σ replaced by $r\sigma/r_0$, which will be useful for later purposes. With this change we substitute (23) into (26) and get

$$\int_\infty^B G(A, t_2) \exp\left[-\frac{r_0^2}{r^2}\frac{A}{2\sigma^2} + \frac{r_0}{\sigma} t_2\right] dt_2$$
$$= \exp\left[-\frac{r_0^2}{r^2}\frac{A}{2\sigma^2}\right] \int_0^A dt_1' \int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \varphi(t_1', t_2')\, dt_2' \int_\infty^B (A - t_1')^{-1/2}$$
$$\cdot \exp\left[-\frac{r_0^2}{2}\frac{(t_2 - t_2')^2}{A - t_1'} + \frac{r_0}{\sigma} t_2\right] dt_2$$
$$= \frac{\sqrt{2\pi}}{r_0} \exp\left[\left(1 - \frac{r_0^2}{r^2}\right)\frac{A}{2\sigma^2}\right] \int_0^A dt_1' \int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \varphi(t_1', t_2')$$
$$\cdot \exp\left[-\frac{1}{2\sigma^2} t_1' + \frac{r_0}{\sigma} t_2'\right] dt_2' \int_\infty^B \frac{r_0}{\sqrt{2\pi}} (A - t_1')^{-1/2}$$
$$\cdot \exp\left[-\frac{r_0^2}{2}\frac{(t_2 - (A - t_1')/\sigma r_0 - t_2')^2}{A - t_1'}\right] dt_2$$

The integral over t_2 can be written

$$\frac{1}{\sqrt{2\pi}} \int_\infty^{B'} \exp\left[-\frac{1}{2} z^2\right] dz,$$

in which

$$B' = r_0(A - t_1')^{-1/2}(B - (A - t_1')/\sigma r_0 - t_2').$$

As $B \to \infty$, $B' \to \infty$ and the integral converges monotonically increasing to 1. The integration over t_2' and t_1' can be considered as a double integral of the form $\iint f_B(t_1', t_2')\, dt_1'\, dt_2'$, in which f_B is bounded in absolute value by

$$|\varphi(t_1', t_2')| \exp\left[-\frac{1}{2\sigma^2} t_1' + \frac{r_0}{\sigma} t_2'\right]$$

which is integrable. Applying the Lebesgue bounded (dominated) convergence theorem, we may take the limit as $B \to \infty$ under the integral. We have

$$\lim_{B\to\infty} f_B(t_1', t_2') = \int_0^A dt_1' \int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \varphi(t_1', t_2') \exp\left[-\frac{1}{2\sigma^2} t_1' + \frac{r_0}{\sigma} t_2'\right] dt_2'.$$

By (32) we may replace the integral on the right in the above equation by minus the integral with same integrand but the t_1' integration running from A to ∞. Thus we get

$$\lim_{B\to\infty} \int_{-\infty}^{B} G(A, t_2) \exp\left[-\frac{r_0^2}{r^2}\frac{A}{2\sigma^2} + \frac{r_0}{\sigma} t_2\right] dt_2 = \xi(A)$$

with

$$\xi(A) = -\frac{\sqrt{2\pi}}{r_0} \exp\left[\left(1 - \frac{r_0^2}{r^2}\right)\frac{A}{2\sigma^2}\right] \int_A^\infty dt_1' \int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \varphi(t_1', t_2')$$

$$\cdot \exp\left[-\frac{1}{2\sigma^2} t_1' + \frac{r_0}{\sigma} t_2'\right] dt_2'.$$

Since ϕ is bounded, we see by (22) that $\varphi(t_1', t_2')$ is bounded in absolute value by const. $h(t_1', t_2')$, which is bounded by const $t_1'^{(n-2)-1}$. In the integration over t_2' we have that

$$\int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \exp\left[\frac{r_0}{\sigma} t_2'\right] dt_2'$$

is bounded by $2\sqrt{t_1'} \exp\left[(r_0/\sigma)\sqrt{t_1'}\right]$ Thus

$$|\xi(A)| < \text{const.} \exp\left[\left(1 - \frac{r_0^2}{r^2}\right)\frac{A}{2\sigma^2}\right] \int_A^\infty t_1'^{(n-1)/2} \exp\left[-\frac{1}{2\sigma^2} t_1' + \frac{r_0}{\sigma}\sqrt{t_1'}\right] dt_1'.$$

We make the substitutions $t_1' = \sigma^2(u + r_0)^2$, $A = \sigma^2 K^2$, then A and K go to ∞ together. Put $\xi(A) = \eta(K)$, then

$$|\eta(K)| < \text{const.} \exp\left[\left(1 - \frac{r_0^2}{r^2}\right)\frac{K^2}{2}\right] \int_{K-r_0}^\infty (u + r_0)^n \exp\left[-\tfrac{1}{2}u^2\right] du.$$

In the integrand, $(u + r_0)^n$ can be bounded by const. u^n, and by partial integration one finds that

$$\int_{K-r_0}^\infty u^n \exp\left[-\tfrac{1}{2}u^2\right] du$$

is bounded by const. $K^{n-1} \exp\left[-\frac{1}{2}(K - r_0)^2\right]$. This leads then to

$$|\eta(K)| < \text{const.} K^{n-1} \exp\left[-\frac{r_0^2}{r^2}\frac{K^2}{2} + r_0 K\right]$$

which $\to 0$ as $K \to \infty$. Q.E.D.

Of the integrated terms (27) and (28) we shall only treat (28), since (27) is a little simpler and follows the same pattern. It can be shown that (23) can be differentiated partially with respect to t_2 under the integral sign, provided $t_2^2 > t_1$. Substituting the result into (28) we obtain, apart from a multiplicative constant,

$$\text{(33)}\quad \iiint (t_1 - t_1')^{-3/2}(B - t_2')\varphi(t_1', t_2') \cdot \exp\left[-\frac{1}{2\sigma^2}t_1 + \frac{r}{\sigma}B - \frac{r_0^2}{2}\frac{(B - t_2')^2}{t_1 - t_1'}\right] dt_2'\, dt_1'\, dt_1$$

in which the integration is over the region $t_2'^2 \leqq t_1' \leqq t_1 \leqq A$. We shall show that (33) $\to 0$ as $B \to \infty$ for fixed A, after which taking the limit $A \to \infty$ yields then trivially 0. Clearly the integrand in (33) approaches 0 as $B \to \infty$. It suffices therefore to show that limit and integral may be interchanged. By the Lebesgue bounded convergence theorem it is sufficient to show that the integrand is bounded in absolute value by an integrable function independent of B (but possibly dependent on A). Let $B_0 > \sqrt{A}$ and consider only values of $B \geqq B_0$. The integrand is bounded in absolute value by $|\varphi(t_1', t_2')| f_1 f_2 f_3$, in which

$$f_1 = \exp\left[\frac{r}{\sigma}B - \frac{r_0^2}{4}\frac{(B - t_2')^2}{t_1 - t_1'}\right], \qquad f_2 = \exp\left[-\frac{r_0^2}{4}\frac{(B - t_2')^2}{t_1 - t_1'}\right]\left(\frac{B - t_2'}{\sqrt{t_1 - t_1'}}\right)^3,$$

and $f_3 = (B - t_2')^{-2}$. Now f_1 is bounded by

$$\exp\left[\frac{r}{\sigma}B - \frac{r_0^2}{4}\frac{(B - \sqrt{A})^2}{A}\right]$$

which is bounded by a constant; f_2 is of the form $y^3 \exp[-(r_0^2/4)y^2]$ and is therefore also bounded by a constant; f_3 is bounded by the constant $(B_0 - \sqrt{A})^{-2}$. Finally we have then that the integrand in (33) is bounded in absolute value by const. $|\varphi(t_1', t_2')|$, which is integrable over the bounded region $t_2'^2 \leqq t_1' \leqq t_1 \leqq A$. Q.E.D.

Appendix 3. We will show that if ϕ is invariant, then $G = 0$ in the region $t_2 \geqq \sqrt{t_1}$. Let $y = t_2/\sqrt{t_1}$, and $y' = t_2'/\sqrt{t_1'}$. If ϕ is invariant, it is a function of y only. Put $\phi(t_1, t_2) = \phi^*(y)$, so that by (22) and (18)

$$\varphi(t_1, t_2) = \text{const. } t_1^{(n/2)-1}(1 - y^2)^{(n/2)-1}(\phi^*(y) - \alpha).$$

After substitution into (23) and making the change of variable $\tau = t_1'/t_1$, we can write (23) as

$$\text{(34)}\quad G(t_1, t_2) = \text{const. } t_1^{n/2} \exp\left[-\frac{r_0^2}{2}y^2\right] \cdot \int_0^1 (1 - y'^2)^{(n/2)-1}(\phi^*(y') - \alpha) f(y, y')\, dy',$$

in which

$$\text{(35)}\quad f(y, y') = \int_0^1 \tau^{(n-1)/2}(1 - \tau)^{-1/2} \exp\left[-\frac{r_0^2}{2}\frac{y^2\tau - 2yy'\sqrt{\tau} + y'^2\tau}{1 - \tau}\right] d\tau$$

Throughout we restrict y and y' to $y > 1$, $y' \leqq 1$. Let the differential operator D_y be defined as

(36)
$$D_y = \frac{1}{r_0^2}\frac{\partial^2}{\partial y^2} - y\frac{\partial}{\partial y} - (n+1)$$

and the operator $D_{y'}$ similarly by replacing in (36) y by y'. Then f satisfies the two equations

(37)
$$D_y f(y, y') = 0$$

(38)
$$D_{y'} f(y, y') = 0$$

Furthermore, it can be seen from (35) that $f \to 0$ if $y \to \infty$ for fixed y', or if $y' \to -\infty$ for fixed y. Two linearly independent solutions of the equation

(39)
$$D_y u(y) = 0$$

are u_1 and u_2, with $u_2(y) = u_1(-y)$, and

(40)
$$u_1(y) = \int_0^\infty t^{(n-1)/2} \exp\left[-\tfrac{1}{2}t + r_0\sqrt{t}\, y\right] dt$$

When $y \to \infty$, $u_1(y) \to \infty$ whereas $u_2(y) \to 0$, with the opposite behavior as $y \to -\infty$. It follows from (37) and (38), from the behavior of the functions u_1 and u_2, and from the behavior of f as $y \to \infty$ or $y' \to -\infty$, that f must equal

(41)
$$f(y, y') = \text{const.}\ u_1(y')u_2(y)$$

Substituting (41) into (34), it remains to be shown that the integral

(42)
$$\int_0^1 (1 - y'^2)^{(n/2)-1}(\phi^*(y') - \alpha)u_1(y')\, dy'$$

equals 0, with u_1 given by (40). Replacing in (40) t by t_1 and in (42) y' by $t_2/\sqrt{t_1}$, the integral (42) is nothing else but the expectation of $\phi - \alpha$ with respect to the distribution specified by $r = r_0$, $\sigma = 1$. Since ϕ is similar of size α this expectation vanishes, Q.E.D.

A METHOD OF GENERATING BEST ASYMPTOTICALLY NORMAL ESTIMATES WITH APPLICATION TO THE ESTIMATION OF BACTERIAL DENSITIES[1]

By Thomas S. Ferguson[2]

University of California, Berkeley

0. Summary. Various minimum χ^2 methods used for generating B.A.N. estimates are summarized, and a new method which generates B.A.N. estimates as roots of certain linear forms is introduced and investigated. As a particular application of the method, the estimation of the bacterial density in an experiment using dilution series is considered.

1. Introduction. The purpose of the present paper is to describe a simple method by which estimates having the usual asymptotic properties of Best Asymptotically Normal (B.A.N.) estimates can be obtained.

Originally B.A.N. estimates were introduced by J. Neyman [1] for a situation in which the underlying probability distributions have a multinomial-like structure. This was followed by a paper by E. W. Barankin and J. Gurland [2] who extended the class of estimation problems for which B.A.N. estimates could be used and also described quite general methods of generation of such estimates. Other results in this direction have been obtained by C. L. Chiang [3] and L. Le Cam [4] and W. Taylor [5].

A best asymptotically normal estimate θ^* of a parameter θ is, loosely speaking, one which is asymptotically normally distributed about the true parameter value, and which is best in the sense that out of all such asymptotically normal estimates it has the least possible asymptotic variance. Thus a B.A.N. estimate will be asymptotically the "most accurate" estimate of a parameter; but the value to a statistician of obtaining such estimates is even greater than is indicated by this. In the aforementioned paper of Neyman, a simple method of testing hypotheses is described which is asymptotically equivalent to the likelihood ratio test and involves the use of the χ^2 statistic and a B.A.N. estimate. It usually turns out that the hardest work in applying this technique is in computing the estimate. Thus it is important to have a number of different methods for computing B.A.N. estimates available to the applied statistician. The usual methods of obtaining B.A.N. estimates will be summarized briefly in section 2.

The objective of all these methods is at least in part a practical one and is essentially two-fold. First, it is hoped that some of these estimates will be easily computable. Second, even though all these estimates have the same

Received July 11, 1956; revised May 7, 1958.

[1] This paper was prepared with the support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171.

[2] Now at the University of California at Los Angeles.

asymptotic properties, they may differ widely in their small sample properties, and it seems reasonable that the choice of the proper estimate to use should depend in part on the behavior of the estimate for small samples. As a consequence, a large class of estimates with best asymptotic properties is proposed with the hope that some of the easily computable estimates will have small sample properties which are reasonably good Blind adherence to the principle of maximum likelihood, for example, may lead to more difficult computations and still yield less accurate estimates than other methods of estimation.

A new approach to generating B.A.N. estimates as roots of linear forms of certain variables is suggested in section 3 In cases where minimum distance methods are applicable, the procedure proposed here leads to estimates which are solutions of equations obtained by simplifying in a suitable manner the equations obtained by the original methods. By way of an example, section 4 contains an application of this approach to the problem of estimation of bacterial density.

2. A review of the minimum χ^2 methods of generating B.A.N. estimates. Since the following methods are to be found in the literature at various levels of generalities, a complete mathematical description of the hypotheses necessary for their validity will be omitted.

Let $X_1, X_2, \cdots, X_n, \cdots$ be a sequence of independent identically distributed s-dimensional random vectors whose distribution depends upon a parameter θ belonging to an open subset Θ of k-dimensional Euclidean space R_k with $k \leq s$. Let $P(\theta) = E(X \mid \theta)$ be the s-dimensional vector of the expectations of the vector X_n, and let $\Sigma(\theta) = var(X \mid \theta) = E\{[X - P(\theta)][X - P(\theta)]'\}$ be the $s \times s$ covariance matrix which is assumed to be finite and non-singular for each $\theta \varepsilon \Theta$. Furthermore, it is assumed that $P(\theta)$ is a one-to-one bicontinuous map from Θ to a subset of s-dimensional Euclidean space with continuous partial derivatives of the second order. Let Z_n be the s-dimensional random vector defined by $nZ_n = \sum_{i=1}^n X_i$.

The quadratic form

$$n[Z_n - P(\theta)]' \Sigma(\theta)^{-1} [Z_n - P(\theta)] \tag{2.1}$$

will be designated by the name of χ^2. The value $\hat{\theta}(Z_n)$ of θ which minimizes this quadratic form will be called the minimum χ^2 estimate of θ. An as example take the multinomial case where there are n independent trials each capable of producing any of $s + 1$ possible outcomes. Let the probability on each successive trial be $p_i(\theta)$ of producing the ith outcome. Let z_i denote the proportion of the trials which result in the ith outcome. Then

$$\chi^2 = n \sum_{i=1}^{s+1} \frac{(z_i - p_i(\theta))^2}{p_i(\theta)} \tag{2.2}$$

is the familiar Pearson χ^2. It may be shown that (2.2) is algebraically equal to the χ^2 of the form (2.1) where the vector Z_n is the vector of the first s z_i's. The χ^2...

antage of (2.1) lies in the fact that it describes a method for estimating pa-
ımeters of continuous distributions.

Barankin and Gurland [2] have shown that the minimum χ^2 estimate, as de-
ned above, is B.A.N. where the X_n have distributions belonging to a Koop-
ıan's family, and Z_n is a vector of sufficient statistics. When the distributions
nder consideration do not form a Koopman's family with sufficient statistics
$_n$, the term B.A.N. estimate is perhaps not entirely justifiable but will be
etained for convenience. The precise definition of B.A.N. estimate to
e adopted is somewhat irrelevant, because the methods reviewed in this sec-
on and the method developed in section 3, give estimates which have the same
symptotic behavior as the minimum χ^2 estimates. In section 3.3, the sense in
hich the estimates are best is stated more precisely.

Starting with this basic minimum χ^2 estimate, several methods may be used
generate large classes of estimates. These methods will be described below.
Iethod I is due essentially to Karl Pearson. Method II as a general method
ıay be found in Barankin and Gurland [2] and Taylor [5], but special cases
ere used earlier (see Berkson [6]). Method III evolved from practical work
nd is of unknown authorship. Method IV is due to Neyman [1].

Method I. Modification. Let $M_n(Z_n, \theta)$ be an $s \times s$ symmetric positive
efinite matrix. The quadratic form

$$Q_n(\theta) = n[Z_n - P(\theta)]' M_n(Z_n, \theta)[Z_n - P(\theta)] \tag{2.3}$$

ill be called the modified or reduced χ^2. The estimate $\hat{\theta}_M(Z_n)$ which minimizes
ıe modified χ^2 with the function $M_n(Z_n, \theta)$ depending only on Z_n and not on
or n, will be called the minimum modified χ^2 estimate of θ. For example, the
stimate which minimizes the Pearson modified χ^2,

$$\chi^2_M = n \sum_{i=1}^{s+1} \frac{(z_i - P_i(\theta))^2}{z_i} \tag{2.4}$$

such an estimate.

Under the condition that $M_n(Z_n, \theta) \to \Sigma^{-1}(\theta)$ in probability as $n \to \infty$ when
is the true value of the parameter, and under certain regularity conditions,
ıe minimum modified χ^2 estimate of θ will have the same asymptotic proper-
es as the minimum χ^2 estimate of θ (or simply $\theta_M(Z_n)$ will be B.A.N., accord-
ıg to the conventions made.)

Method II. Transformation. Let $g(x)$ be any function from R_s to R_s with
ontinuous first partial derivatives

$$g(x) = \begin{pmatrix} g_1(x_1, \cdots, x_s) \\ \vdots \\ g_s(x_1, \cdots, x_s) \end{pmatrix} \tag{2.5}$$

Let the $s \times s$ matrix of first partial derivatives be denoted by

$$(2.6) \qquad g(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} g_1 & \cdots & \frac{\partial}{\partial x_1} g_s \\ \vdots & & \\ \frac{\partial}{\partial x_s} g_1 & \cdots & \frac{\partial}{\partial x_s} g_s \end{pmatrix}$$

We shall call the quadratic form

$$(2.7) \qquad n[g(Z_n) - g(P(\theta))]'[g(P(\theta))\Sigma(\theta)g(P(\theta))']^{-1}[g(Z_n) - g(P(\theta))]$$

the transformed χ^2. More generally, we may consider the combinations of Methods I and II, and replace the matrix of the quadratic form (2.7) by an estimate

$$(2.8) \qquad Q_n(\theta) = n[g(Z_n) - g(P(\theta))]'M_n(Z_n, \theta)[g(Z_n) - g(P(\theta))].$$

We assume that $M_n(Z_n, \theta) \to [g(P(\theta))\Sigma(\theta)g(P(\theta))']^{-1}$ in probability and the regularity conditions needed for Method I. In addition, one needs regularity conditions on g, namely that g is a one-to-one bicontinuous map from a neighborhood of $P(\Theta)$ into R_s, with continuous partial derivatives of the second order and that the matrix $g(P(\theta))$ is nonsingular for each $\theta \varepsilon \Theta$. Then the minimum transformed χ^2 estimates, that is the value $\hat{\theta}_T(Z_n)$ of θ minimizing (2.7), will be a B.A.N. estimate of θ.

This method of generating B.A.N. estimates also applies to the χ^2 of (2.2); for example, letting $g_i(x)$ be the real-valued transformation applied to the ith cell

$$(2.9) \qquad \chi^2 = n \sum_{i=1}^{s+1} \frac{[g_i(z_i) - g_i(p_i(\theta))]^2}{p_i(\theta)g_i'(p_i(\theta))}$$

or modified,

$$(2.10) \qquad \chi^2 = n \sum_{i=1}^{s+1} \frac{[g_i(z_i) - g_i(p_i(\theta))]^2}{z_i\, g_i'(z_i)^2}.$$

The well-known example of Berkson [6] is of the type (2.10).

Many times the functions g_i may be chosen so that $g_i(p_i(\theta))$ is a linear function of the parameters $\theta_1, \cdots, \theta_k$. In such cases finding the value of θ which minimizes the χ^2 of equation (2.10) results in solving k linear equations in k unknowns. The reader may consult the paper of W. Taylor [5] for examples.

Method III. Expansion in a Taylor series about a $O(\sqrt{n})$-consistent estimate. An estimate θ_n^* of θ will be called $O(\sqrt{n})$-consistent if $\sqrt{n}(\theta_n^* - \theta)$ is bounded in probability uniformly in n when θ is the true value of the parameter; that is, for every $\epsilon > 0$ and $\theta \varepsilon \Theta$, there exists a number B so large that for every $n = 1, 2, \cdots$

(2.11) $$P[\sqrt{n}\mid\theta_n^*-\theta\mid>B\mid\theta]<\epsilon.$$

Many types of estimates satisfy this requirement. For example, under certain regularity conditions, estimation by the method of moments yields estimates θ_n^* for which $\sqrt{n}(\theta_n^*-\theta)$ is asymptotically normal when θ is the true value of the parameter. This follows from a theorem of Cramér [7], p. 366, which states that certain functions of the moments are asymptotically normal. Such asymptotically normal estimates as this are obviously $O(\sqrt{n})$-consistent.

One may try to apply a correction to θ_n^* by an application of the method of expansion in a Taylor series to get an estimate closer to the minimum χ^2 estimate. It is known, however, that one such application to a $O(\sqrt{n})$-consistent estimate will give a B.A.N. estimate. More specifically, consider the expansion of some one of the previously mentioned χ^2's (modified and/or transformed) in a Taylor series to the second degree terms about a $O(\sqrt{n}$-consistent estimate θ_n^* of θ.

(2.12) $$\chi^2(\theta)=\chi^2(\theta_n^*)+\dot{\chi}^2(\theta_n^*)'(\theta-\theta_n)+\tfrac{1}{2}(\theta-\theta_n^*)'\ddot{\chi}^2(\theta_n^*)(\theta-\theta_n^*)+\text{Rem.}$$

where $\dot{\chi}^2(\theta)$ is the $k\times 1$ vector of first derivatives of $\chi^2(\theta)$ and $\ddot{\chi}^2(\theta)$ is the $k\times k$ matrix of second derivatives of $\chi^2(\theta)$.

(2.13) $$\dot{\chi}^2(\theta)=\begin{pmatrix}\dfrac{\partial}{\partial\theta_1}\chi^2(\theta)\\ \vdots\\ \dfrac{\partial}{\partial\theta_k}\chi^2(\theta)\end{pmatrix}$$

(2.14) $$\ddot{\chi}^2(\theta)=\begin{pmatrix}\dfrac{\partial^2}{\partial\theta_1^2}\chi^2(\theta) & \cdots & \dfrac{\partial^2}{\partial\theta_1\,\partial\theta_k}\chi^2(\theta)\\ \vdots & & \\ \dfrac{\partial^2}{\partial\theta_k\,\partial\theta_1}\chi^2(\theta) & \cdots & \dfrac{\partial^2}{\partial\theta_k^2}\chi^2(\theta)\end{pmatrix}.$$

Instead of finding that value of θ which minimizes $\chi^2(\theta)$, one may discard the remainder term and find that value $\hat{\theta}_n$ of θ which minimizes the first three terms of the expansion. This estimate $\hat{\theta}_n$ will then be a B.A.N. estimate of θ. This method of generating B.A.N. estimates is important because it leads to k linear equations in k unknowns and is thus comparatively easy to apply.

Method IV. Linearization of the side conditions. This method, due to Neyman [1], was proposed with the specific intention of finding a B.A.N. estimate which could be computed by solving linear equations. In minimizing some χ^2 like (2.1), one may consider the vector P as the vector of parameters which are subject to certain restrictions, called side conditions, due to the dependence of P on θ. If there are s independent components of the vector P and k parameters, there will be $s-k$ side conditions on the p's.

(2.16) $$f_j(p_1,\cdots,p_s)=0\qquad\text{for}\quad j=1,\cdots,k-s$$

ne may then minimize χ^2 subject to these side conditions by the method of agrange multipliers. However, a simpler procedure would be to minimize χ^2 ubject to the linearized counterpart of (2.15), that is, the first two terms of ne Taylor series expansion about the point z_n. The solution for the estimate nen only requires solution of linear equations. For a fuller account of the subect, the reader should consult the papers of Neyman and of Barankin and urland. The outline of the method given here is added only for the sake of ompleteness and no mention of the method will be made in the later sections f the paper.

3. B.A.N. estimates as roots of linear forms. The method customarily used to nd a minimum χ^2 estimate is to differentiate χ^2 with respect to each of the arameters separately, set the results equal to zero and solve the resulting system f simultaneous equations. For example, one may differentiate the χ^2 of the quation (2.4) and obtain the equations

$$-2n \sum_{i=1}^{s} \frac{z_i - p_i(\theta)}{z_i} \frac{\partial p_i(\theta)}{\partial \theta_j} = 0 \qquad \text{for } j = 1, 2, \cdots, k, \tag{3.1}$$

r one may differentiate the χ^2 of equation (2.3) with $M_n(Z_n, \theta)$ a function of Z_n only, such that $M_n(Z_n) \to \Gamma(\theta)$ in probability and the regularity conditions old, and obtain

$$-n\dot{P}(\theta)M(Z_n)(Z_n - P(\theta)) = 0 \tag{3.2}$$

here $\dot{P}(\theta)$ is the $k \times s$ matrix of first partial derivatives of the vector $P(\theta)$,

$$\dot{P}(\theta) = \begin{pmatrix} \frac{\partial P_1(\theta)}{\partial \theta_1} & \cdots & \frac{\partial P_s(\theta)}{\partial \theta_1} \\ \vdots & & \\ \frac{\partial P_1(\theta)}{\partial \theta_k} & \cdots & \frac{\partial P_s(\theta)}{\partial \theta_k} \end{pmatrix} \tag{3.3}$$

nd the 0 is the $k \times 1$ vector with a zero term in each component so that (3.2) epresents k equations in k unknowns.

Well-chosen roots to equations such as (3.1) and (3.2) are B.A.N. estimates f the unknown parameters. This suggests that instead of starting with a quadatic form in $(Z_n - P(\theta))$ and finding values of θ which make the form a mininum, it may be simpler to start with an arbitrary linear form in $(Z_n - P(\theta))$ nd find the roots. Roots of certain such linear forms, namely, (3.1) and (3.2), vill be B.A.N. estimates. Furthermore, such a method of generating B.A.N. stimates will probably satisfy the requirement that they be easy to compute. t is the purpose of this section to investigate the asymptotic distribution of oots of linear forms in $(Z_n - P(\theta))$, and the conditions for such roots to be B.A.N. estimates of the parameters.

3.1. Preliminary lemma. This section contains an implicit function theorem needed for the proof of the main theorem. First an implicit function theorem

which can be found in Pierpont [8], p. 293, for example is stated, from which the lemma of this section will follow. The unicity of the implicit function is stated in a somewhat stronger form than found in Pierpont. This strengthening can be obtained by modifying his proof slightly and the details of the proof need not be given here.

Let $F(x, u)$ be a function of variables $x \varepsilon R_s$ and $u \varepsilon R_k$ with values in R_k. Let $a \varepsilon R_s$ and $b \varepsilon R_k$, and assume that

(i) *$F(x, u)$ is continuous and $F_u(x, u)$ exists and is continuous in a neighborhood of the point (a, b).*

(ii) *$F(a, b) = 0$ and $F_u(a, b)$ is nonsingular. Then, there exists a neighborhood N of a, and a function $\phi(x)$ from R_s to R_k, such that*

(1) *$\phi(x)$ is continuous in N,*

(2) *$\phi(a) = b$,*

(3) *$F(x, \phi(x)) = 0$ for $x \varepsilon N$, and*

(4) *(uniqueness) there exists a neighborhood N' of the point b such that for $u \varepsilon N'$ and $x \varepsilon N$, $F(x, a) \neq 0$ unless $u = \phi(x)$.*

In the above theorem $F_u(x, u)$ represents the $k \times k$ matrix of partial derivatives of $F(x, u)$ with respect to u, as in equation (3.3). The assumption of continuity of $F_u(x, u)$ means that each component of the matrix is assumed to be continuous.

The following lemma is an extension of this theorem, similar to that found in Graves [9], p. 144, to the situation in which $F(x, u)$ is known to vanish along some curve in R_{s+k}, rather than just at one point.

LEMMA. *Let $F(x, u)$ be a function of variables $x \varepsilon R_s$ and $u \varepsilon R_k$ with values in R_k, $k \leqq s$. Let $p(u)$ be a function from some set $D \subset R_k$ to R_s, and assume that*

(i) *D is an open set,*

(ii) *$p(u)$ is one-to-one and inversely continuous from D into R_s,*

(iii) *there is a neighborhood of the curve $\{(p(u), u): u \varepsilon D\}$ in which $F(x, u)$ is continuous and $F_u(x, u)$ exists and is continuous.*

(iv) *$F(p(u), u) = 0$ and $F_u(p(u), u)$ is nonsingular for every $u \varepsilon D$.*

Then, there exists a neighborhood N of the set $S = \{p(u): u \varepsilon D\}$ and a function $\phi(x)$ from R_s to R_k such that

(a) *$\phi(x)$ is continuous in N,*

(b) *$\phi(p(u)) = u$ for $u \varepsilon D$,*

(c) *$F(x, \phi(x)) = 0$ for $x \varepsilon N$,* and

(d) *there exists a neighborhood of the curve $\{(p(u), u): u \varepsilon D\}$ in which the only zeros of the function $F(x, u)$ are the points $(x, \phi(x))$.*

PROOF. From the previous implicit function theorem, for every $u \varepsilon D$, there is a neighborhood $N_{p(u)}$ of the point $p(u)$ and a function $\phi_u(x)$ from R_s to R_k such that

(1) $\phi_u(x)$ is continuous in $N_{p(u)}$,

(2) $\phi_u(p(u)) = u$,

(3) $F(x, \phi_u(x)) = 0$ for $x \varepsilon N_{p(u)}$, and

(4) for y in some neighborhood N_u of the point u, and $x \varepsilon N_{p(u)}$

$$F(x, y) \neq 0 \quad \text{unless} \quad y = \phi_u(x).$$

Using the inverse continuity of the function $p(u)$, and the continuity of the function $\phi_u(x)$, we may replace the neighborhoods $N_p(u)$ by spherical neighborhoods $N'_{p(u)}$ with the two additional properties that

(5) if $p(u_1) \varepsilon N'_{p(u_2)}$ for some u_1 and $u_2 \varepsilon D$, then $u_1 \varepsilon N_{u_2}$

and

(6) if $x \varepsilon N'_{p(u)}$ for some $u \varepsilon D$, then $\phi_u(x) \varepsilon N_u$.

Now consider spherical neighborhoods $N''_{p(u)}$ with radii equal to ⅓ that of $N'_{p(u)}$, and let N denote $\bigcup_{u \varepsilon D} N''_{p(u)}$ The set N is then obviously a neighborhood of the set S.

We will show that if $x_0 \varepsilon N''_{p(u_1)} \cap N''_{p(u_2)}$, then $\phi_{u_1}(x_0) = \phi_{u_2}(x_0)$. For since $N''_{p(u_1)} \cap N''_{p(u_2)}$ is not empty, either $p(u_1) \varepsilon N'_{p(u_2)}$ or $p(u_2) \varepsilon N'_{p(u_1)}$. Suppose without loss of generality that the former is true; then since $u_1 \varepsilon N_{u_2}$ and

$$F(p(u_1), \phi_{u_2}(p(u_1))) = 0,$$

we have $\phi_{u_2}(p(u_1)) = u_1$. Furthermore, for $x \varepsilon N''_{p(u_1)} \cap N'_{p(u_2)}$, $\phi_{u_2}(x)$ is continuous and satisfies $F(x, \phi_{u_2}(x)) = 0$, but $\phi_{u_1}(x) \varepsilon N_{u_1}$ for $x \varepsilon N'_{p(u_1)}$ and thus $\phi_{u_1}(x)$ is the unique function, continuous in $N''_{p(u_1)}$ and such that

$$\phi_{u_1}(p(u_1)) = u_1 \quad \text{and} \quad F(x, \phi_{u_1}(x)) = 0.$$

Hence, $\phi_{u_1}(x_0) = \phi_{u_2}(x_0)$.

Thus for $x \varepsilon N$ we may define $\phi(x) = \phi_u(x)$ for any u for which $x \varepsilon N''_{p(u)}$, since such a definition is unique. Now parts (a), (b), and (c) of the conclusion of the lemma are obvious. As for (d), the neighborhood can be chosen to be $\bigcup_{u \varepsilon D}[N''_{p(u)} \times N_u]$.

3.2. The main theorem. Let Z_n, $n = 1, 2, \cdots$ be a sequence of s-dimensional random vectors whose distribution depends upon a parameter θ in some set $\Theta \subset R_k$, $k \leqq s$. Let $P(\theta)$ be a function from Θ to R_s.

ASSUMPTION 1. *Θ is an open set.*

ASSUMPTION 2. *$\mathcal{L}\{\sqrt{n}(Z_n - P(\theta)) \mid \theta\} \to \mathcal{L}(Z)$ where Z is a normal random vector with mean zero and variance-covariance matrix $\Sigma(\theta)$. (That is,*

$$EZ = 0, \quad EZZ' = \Sigma(\theta).)$$

The convergence used above is convergence in law or in distribution. Assumption 2 states that when θ is the true value of the parameter, the distribution of $\sqrt{n}(Z_n - P(\theta))$ converges to a normal distribution with mean zero and variance-covariance matrix $\Sigma(\theta)$. The law degenerate at some point a will be denoted by $\mathcal{L}(a)$. Thus $\mathcal{L}(X_n) \to \mathcal{L}(a)$ means that X_n converges in probability to a.

ASSUMPTION 3. *The mapping $P(\theta)$ from Θ into R_s is homeomorphic (that is, one-to-one and bicontinuous) and continuously differentiable.*

Let $f(x, \theta)$ be a $k \times s$ matrix for each $x \varepsilon R_s$ and $\theta \varepsilon \Theta$.

ASSUMPTION 4. *There is a neighborhood $N_0 \subset R_s \times \Theta$ of the set*

$$\{(P(\theta), \theta) : \theta \varepsilon \Theta\}$$

within which $f(x, \theta)$ *and* $\partial/\partial\theta_j f(x, \theta)$ *for* $j = 1, 2, \cdots, k$ *are continuous jointly in* (x, θ).

Let $b(\theta) = f(P(\theta), \theta)$ and let $\dot{P}(\theta)$ be the $k \times s$ matrix of partial derivatives of $P(\theta)$, given by equation (3.3).

ASSUMPTION 5. *The matrix* $\dot{P}(\theta)b(\theta)'$ *is nonsingular for each* $\theta \in \Theta$.

Let

$$F(x, \theta) = f(x, \theta)(x - P(\theta)). \tag{3.4}$$

This is the linear form which will be used in the sequel to generate B.A.N. estimates of the parameter θ. The following theorem shows immediately that the root to the equation $F(Z_n, \theta) = 0$ will be a $O(\sqrt{n})$-consistent estimate of θ.

THEOREM 1. *Under assumptions* 1 *through* 5, *there exists a neighborhood* N *of the set* $S = \{P(\theta): \theta \in \Theta\}$ *and a unique function* $\hat{\theta}(x)$ *from* R_s *to* R_k *continuous in* N, *such that* $\hat{\theta}(P(\theta)) = \theta$ *for* $\theta \in \Theta$, *and* $F(x, \hat{\theta}(x)) = 0$ *for* $x \in N$. *Moreover,* $\mathcal{L}\{\sqrt{n}(\hat{\theta}(Z_n) - \theta) | \theta\} \rightarrow \mathcal{L}(Y)$ *where* Y *is a normal random vector with mean zero and variance-covariance matrix given by*

$$[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)\Sigma(\theta)b(\theta)'[\dot{P}(\theta)b(\theta)']^{-1}$$

PROOF. $F(P(\theta), \theta) = 0$ and

$$F_\theta(x, \theta) = f_\theta(x, \theta)(x - P(\theta)) - \dot{P}(\theta)f(x, \theta)' \tag{3.5}$$

where $f_\theta(x, \theta)$ represents the $k \times k \times s$ cubic matrix of partial derivatives of the $k \times s$ matrix $f(x, \theta)$ with respect to θ. To avoid confusion we will write out the first term of this difference completely. Denote the function in the ith row, jth column of $f(x, \theta)$ by $f_{ij}(x, \theta)$, and let $P_j(\theta)$ and x_j represent the jth component of the vectors $P(\theta)$ and x. Then,

$$f_\theta(x, \theta)(x - P(\theta)) = \sum_{j=1}^{s} \begin{pmatrix} \frac{\partial}{\partial\theta_1} f_{1j} & \cdots & \frac{\partial}{\partial\theta_1} f_{sj} \\ \vdots & & \\ \frac{\partial}{\partial\theta_k} f_{1j} & \cdots & \frac{\partial}{\partial\theta_k} f_{sj} \end{pmatrix} (x_j - P_j(\theta)). \tag{3.6}$$

It is now easily checked that formula (3.5) holds. Hence,

$$F_\theta(P(\theta), \theta) = -\dot{P}(\theta)b(\theta)' \tag{3.7}$$

which, by assumption, is nonsingular for every $\theta \in \Theta$. Thus the hypotheses of the lemma of the previous section are satisfied and the first part of the theorem is proved.

To prove the second part, expand $F(x, \theta)$ about the point $\hat{\theta}(x)$ to one term using the formula

$$F(x, \theta) = F(x, \hat{\theta}(x)) + \left[\int_0^1 F_\theta\{x, \hat{\theta}(x) + \lambda(\theta - \hat{\theta}(x)\} \, d\lambda\right]' (\theta - \hat{\theta}(x)) \tag{3.8}$$

which may easily be checked. By the integral of a matrix we mean the matrix of the integrals of each term separately. For each $\theta \varepsilon \Theta$, formula (3.8) is valid *whenever x is sufficiently close to $p(\theta)$*, so that $(x, \hat{\theta}(x))$ is in a spherical neighborhood of $(p(\theta), \theta)$ contained entirely in N_0. We may replace x by Z_n in (3.8) and multiply both sides by $\sqrt{n}$.

$$\sqrt{n}\left[-\int_0^1 F_\theta\{Z_n, \hat{\theta}(Z_n) + \lambda(\theta - \hat{\theta}(Z_n))\}\, d\lambda\right]' (\hat{\theta}(Z_n) - \theta) = f(Z_n, \theta)\sqrt{n}\,(Z_n - P(\theta)). \tag{3.9}$$

We now invoke the theorems of Slutsky (see [10], section 2, theorem 2, or [4]). From assumption 1, $\mathcal{L}(Z_n \mid \theta) \to \mathcal{L}(P(\theta))$ Hence by Slutsky's theorem, *since $f(x, \theta)$ is continuous in a neighborhood of $(p(\theta), \theta)$*,

$$\mathcal{L}(f(Z_n, \theta) \mid \theta) \to \mathcal{L}(f(P(\theta), \theta)) = \mathcal{L}(b(\theta)). \tag{3.10}$$

Slutsky's theorem also gives

$$\mathcal{L}(f(Z_n, \theta)\sqrt{n}(Z_n - P(\theta)) \mid \theta) \to \mathcal{L}(b(\theta)Z) \tag{3.11}$$

where Z is a normal vector with zero mean and variance-covariance matrix $\Sigma(\theta)$. Since $\mathcal{L}(Z_n \mid \theta) \to \mathcal{L}(P(\theta))$ and $\mathcal{L}(\hat{\theta}(Z_n) \mid \theta) \to \mathcal{L}(\hat{\theta}(P(\theta))) = \mathcal{L}(\theta)$, we may apply the Lebesgue bounded convergence theorem to the integral in (3.9).

$$\mathcal{L}\left\{\int_0^1 F_\theta[Z_n, \hat{\theta}(Z_n) + \lambda(\theta - \hat{\theta}(Z_n))]\, d\lambda \mid \theta\right\} \to \mathcal{L}\left\{\int_0^1 F_\theta[P(\theta), \theta]\, d\lambda\right\} = \mathcal{L}\{F_\theta(P(\theta), \theta)\} = \mathcal{L}\{-\dot{P}(\theta)b(\theta)'\} \tag{3.12}$$

by equation (3.7). Another application of Slutsky's theorem allows us to conclude

$$\mathcal{L}\{\sqrt{n}(\hat{\theta}(Z_n) - \theta) \mid \theta\} \to \mathcal{L}\{[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)Z\}. \tag{3.13}$$

Denoting $[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)Z$ by Y, we see that Y is a normal random vector, with mean zero and covariance matrix

$$\begin{aligned} EYY' &= E[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)ZZ'b(\theta)'[\dot{P}(\theta)b(\theta)']^{-1} \\ &= [b(\theta)\dot{P}(\theta)']^{-1}b(\theta)\Sigma(\theta)b(\theta)'[\dot{P}(\theta)b(\theta)']^{-1}. \end{aligned} \tag{3.14}$$

3.3. Applications. The theorem just proved allows some immediate inferences. The important point to notice in this theorem is that the asymptotic distribution of $\sqrt{n}(\hat{\theta}(Z_n) - \theta)$ depends on the function $f(x, \theta)$ only through its values along the curve $\{(P(\theta), \theta) : \theta \varepsilon \Theta\}$. Thus if the linear form

$$F(Z_n, \theta) = f(Z_n, \theta)(Z_n - P(\theta))$$

has a root which is already a B.A.N. estimate of θ, any linear form

$$g(Z_n, \theta)(Z_n - P(\theta)),$$

in which the function $f(x, \theta)$ is replaced by any function $g(x, \theta)$ satisfying assumption 4 and for which $g(P(\theta), \theta) = f(P(\theta), \theta)$, will have a root which is also a B.A.N. estimate of θ, since the asymptotic distribution of the two roots will be the same.

For example, equation (3.2) (neglecting the factor n which is immaterial as far as roots are concerned) is a linear form of the type $f(Z_n, \theta)(Z_n - P(\theta))$ for which

$$f(Z_n, \theta) = \dot{P}(\theta)M(Z_n). \tag{3.15}$$

Since $M(Z_n)$ converges in probability to $\Sigma(\theta)^{-1}$ when θ is the true value of the parameter, $M(P(\theta)) = \Sigma(\theta)^{-1}$ so that

$$b(\theta) = \dot{P}(\theta)\Sigma(\theta)^{-1} \tag{3.16}$$

Now consider functions

$$f_1(Z_n, \theta) = b(\theta) \quad \text{and} \quad f_2(Z_n, \theta) = L(Z_n)M(Z_n) \tag{3.17}$$

where L is a matrix continuous in a neighborhood of $\{P(\theta):\theta \varepsilon \Theta\}$, such that $L(P(\theta)) = \dot{P}(\theta)$. If $f_1(Z_n, \theta)$ is used, we must also assume that $b(\theta)$ has a continuous derivative. In these circumstances, whenever the root to equation (3.2) is a B.A.N. estimate, roots to the linear forms involving $f_1(Z_n, \theta)$ and $f_2(Z_n, \theta)$ will be B.A.N. also.

Now we will show directly the exact conditions under which there will be a root of a linear form which will be "best" out of the class of all roots of linear forms; that is, the exact conditions under which there is a value of $b(\theta)$ which minimizes the variance (3.14).

Of two n by n matrices, A and B, A will be said to be smaller than B, in symbols $A < B$, if and only if $B - A$ is positive semi-definite; that is, if

$$x'[B - A]x \geqq 0$$

for every n-dimensional vector x. Thus of two unbiased estimates of a vector parameter θ, T_1 and T_2, with covariance matrices respectively A and B, T_1 would be preferred to T_2 if $A < B$, since the unbiased estimate $x'T_1$ of the parameter $x'\theta$ will have a smaller variance than the unbiased estimate $x'T_2$ of the same parameter.

THEOREM 2. *If in addition to assumptions* 1 *through* 5 *there exists an* s *by* s *nonsingular matrix* $\Sigma_0(\theta)$ *such that*

$$\Sigma(\theta)\Sigma_0(\theta)\dot{P}(\theta)' = \dot{P}(\theta)' \tag{3.18}$$

then the asymptotic covariance matrix of $\hat{\theta}(Z_n)$ *taken on its minimum value when* $b(\theta) = \dot{P}(\theta)\Sigma_0(\theta)$. *The minimum value is then* $[\dot{P}(\theta)\Sigma_0(\theta)\dot{P}(\theta)']^{-1}$.

PROOF. For simplicity of notation the θ will be omitted. From assumption 5, $\dot{P}$ is of full rank so that $[\dot{P}\Sigma_0\dot{P}]$ is nonsingular. The inequality

$$(b'[\dot{P}b']^{-1} - \Sigma_0\dot{P}'[\dot{P}\Sigma_0\dot{P}']^{-1})\Sigma(b[\dot{P}b']^{-1} - \Sigma_0\dot{P}'[\dot{P}\Sigma_0\dot{P}]^{-1}) \geqq 0 \tag{3.19}$$

which holds since Σ is positive semi-definite, yields

$$[b\dot{P}']^{-1}b\Sigma b'[\dot{P}b']^{-1} - [\dot{P}\Sigma_0\dot{P}']^{-1} \geqq 0. \tag{3.20}$$

Yet it is easily checked that equality is attained if $b = \dot{P}\Sigma_0'$. qed.

The assumption of the existence of a matrix $\Sigma_0(\theta)$ satisfying (3.18) holds for example when $\Sigma(\theta)$ is nonsingular. Then $b(\theta) = \dot{P}(\theta)\Sigma(\theta)^{-1}$ as was found in equation (3.16). However, in other important cases, for example in the multinomial case with the χ^2 of equation (2.2), the matrix $\Sigma(\theta)$ is singular. The following lemma which may be proved without difficulty, will perhaps be of aid in checking whether a Σ_0 satisfying (3.18) exists at all.

LEMMA. *In order that there exist a nonsingular matrix $\Sigma_0(\theta)$ satisfying* (3.18), *it is necessary and sufficient that the range space of $\dot{P}(\theta)'$ be contained in the range space of $\Sigma(\theta)$: that is, for every vector x there exists a vector $y(\theta)$ such that*

$$\Sigma(\theta)y(\theta) = \dot{P}(\theta)'x.$$

In certain cases one can find the matrix Σ_0 which satisfies (3.18). We shall do it now for the multinomial case. In this case the vector $P(\theta)$ is simply the vector of cell probabilities, and is $s + 1$ dimensional. The matrix $\Sigma(\theta)$ is found to be

$$\Sigma(\theta) = \begin{pmatrix} p_1(\theta) - p_1^2(\theta) & -p_1(\theta)p_2(\theta) & \cdots & -p_1(\theta)p_{s+1}(\theta) \\ -p_1(\theta)p_2(\theta) & p_2(\theta) - p_2^2(\theta) & & \\ \vdots & & & \\ -p(\theta)p_{s+1}(\theta) & & \cdots & p_{s+1}(\theta) - p_{s+1}^2(\theta) \end{pmatrix} \tag{3.21}$$

which may be expressed simply as

$$\Sigma(\theta) = B(\theta) - P(\theta)P(\theta)' \tag{3.22}$$

where $B(\theta)$ is the diagonal matrix

$$B(\theta) = \begin{pmatrix} p_1(\theta) & 0 & \cdots & 0 \\ 0 & p_2(\theta) & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \\ 0 & & & p_{s+1}(\theta) \end{pmatrix} \tag{3.23}$$

Then, as suggested by the χ^2 of (2.2), put $\Sigma_0(\theta) = B(\theta)^{-1}$.

$$\Sigma(\theta)\Sigma_0(\theta)\dot{P}(\theta)' = B(\theta)B(\theta)^{-1}\dot{P}(\theta)' - P(\theta)P(\theta)'B(\theta)^{-1}\dot{P}(\theta)'. \tag{3.24}$$

It is easily seen that

$$P(\theta)'B(\theta)^{-1}\dot{P}(\theta)' = \left(\sum_{i=1}^{s+1} \frac{\partial}{\partial\theta_1} p_i(\theta), \sum_{i=1}^{s+1} \frac{\partial}{\partial\theta_2} p_i(\theta), \cdots, \sum_{i=1}^{s+1} \frac{\partial}{\partial\theta_k} p_i(\theta)\right). \tag{3.25}$$

This vector must be zero since $\sum_{i=1}^{s+1} p_i(\theta) = 1$. Hence, the equality (3.18) is satisfied. Thus applying Theorem 2, roots of the linear form

$$\sum_{i=1}^{s+1} (z_i - p_i(\theta))f_{ij}(s_1, \cdots, s_{s+1}, \theta) = 0 \qquad j = 1, 2, \cdots k, \tag{3.26}$$

will be "best" when $f_{ij}(p_1(\theta), \cdots, p_{s+1}(\theta), \theta) = \partial/\partial\theta_j \log p_i(\theta)$.

It may further be shown in the multinomial case, that if the $f_{ij}(z, \theta)$ are chosen to be independent of z and equal to $\partial/\partial\theta_j \log p_i(\theta)$, equation (3.25) will be the derivative of the log of the likelihood function set equal to zero, so that one has immediately that the maximum likelihood estimate, in addition to the minimum modified χ^2 estimate, is an estimate which is given by the root of a certain linear form. One would expect that the linear form (3.25) in which the functions f_{ij} do not depend on θ at all would be somewhat easier to solve for θ. It is this type of linear form which is suggested in section 4 as a method for estimating the bacterial density in a liquid.

We will now apply the preceding theorem to the various minimum χ^2 methods discussed previously.

Application to the transformed χ^2. The method of generating B.A.N. estimates described in Theorems 1 and 2 also applies easily to the transformed χ^2 of equations (2.8) and (2.10). For example, the derivative of the χ^2 of equation (2.8) with $T(Z_n)$ depending on Z_n only, and not on θ, is found to be

$$\frac{\partial}{\partial\theta}\chi_T^2 = n\dot{P}(\theta)\dot{g}(P(\theta))T(Z_n)(g(Z_n) - g(P(\theta))). \tag{3.27}$$

Assumption 1 of Theorem 1 becomes in this case

$$\mathcal{L}(\sqrt{n}(g(Z_n) - g(P(\theta))) \mid \theta) \to \mathcal{L}(Z) \tag{3.28}$$

where Z is a normal random vector with zero mean and variance-covariance matrix $[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']$. This may easily be checked by expanding $g(Z_n)$ in a Taylor series about the point $P(\theta)$, and invoking asymptotic normality of $\sqrt{n}(Z_n - P(\theta))$. The only requirement on the function $g(x)$ is that it have a continuous derivative in a neighborhood of the curve $\{P(\theta): \theta \in \Theta\}$. If in addition $\dot{g}(P(\theta))$ is nonsingular for each $\theta \in \Theta$, $[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']^{-1}$ will exist and $b(\theta)$ is found to be

$$b(\theta) = \dot{P}(\theta)\dot{g}(P(\theta))[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']^{-1}. \tag{3.29}$$

Thus, if the root to equation (3.27) is a B.A.N. estimate, the root to the linear form

$$f(Z_n, \theta)(g(Z_n) - g(P(\theta))) = 0 \tag{3.30}$$

will also be a B.A.N. estimate, provided that f satisfies Assumption 4, and that $f(P(\theta), \theta) = b(\theta)$.

The linear form corresponding to the transformed multinomial χ^2 of (2.10) may be computed as before. It becomes

$$\sum_{i=1}^{k} [g_i(z_i) - g_i(p_i(\theta))]f_{ij}(z_1, \cdots, z_{k-1}, \theta) = 0 \qquad j = 1, 2, \cdots, r \tag{3.31}$$

where

$$f_{ij}(p_1(\theta), \cdots, p_{k-1}(\theta), \theta) = \left[\frac{\partial}{\partial\theta_j} p_i(\theta)\right]\frac{1}{p_i(\theta)g_i'(p_i(\theta))}. \tag{3.32}$$

Under assumptions 1 through 5, and the assumptions that each $g_i(x)$ is continuous in a neighborhood of the curve $\{x : x = p_i(\theta),\ \theta \varepsilon \Theta\}$ and that

$$g_i'(p_i(\theta)) \neq 0,$$

the roots to equation (3.31) will be B.A.N. estimates of the parameters.

Application to the expansion of χ^2 in a Taylor series. Let θ_n^* be a $O(\sqrt{n})$-consistent estimate of the parameter θ. To find the minimum value of the right hand side of equation (2.12) without the remainder term, we take a derivative and solve for the root $\hat{\theta}$.

$$\hat{\theta}_n = \theta_n^* - \ddot{\chi}^2(\theta_n)^{-1}\dot{\chi}^2(\theta_n) \tag{3.33}$$

If we use the modified χ^2 of equation (2.3) for this procedure with M a function of Z_n only, for example $M(Z_n) = \Sigma(\theta_n^*)^{-1}$, the first two derivatives are

$$\begin{aligned} \dot{\chi}^2(\theta) &= 2n\dot{P}(\theta)\Sigma(\theta_n^*)^{-1}(Z_n - P(\theta)) \\ \ddot{\chi}^2(\theta) &= 2n\dot{P}(\theta)\Sigma(\theta_n^*)^{-1}\dot{P}(\theta)' - 2n\ddot{P}(\theta)\Sigma(\theta_n^*)^{-1}(Z_n - P(\theta)). \end{aligned} \tag{3.34}$$

where $\ddot{P}(\theta)$ is the $k \times k \times s$ cubic matrix of second partial derivatives of the vector $P(\theta)$.

If, on the other hand, we take the linear form with the function $f(Z_n, \theta)$ not depending on θ, say to be $\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}$, and expand it about θ_n^* to the first power and solve for θ, we have

$$\hat{\theta}_n = \theta_n^* + [\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}\dot{P}(\theta_n^*)]^{-1}\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}(Z_n - P(\theta_n^*)). \tag{3.35}$$

If one compares the estimates (3.35) with the estimates (3.33) with equations (3.34) substituted, one sees that the former require less computation, and that by the amount in the second term of the expression for $\ddot{\chi}^2(\theta)$, involving all the second partial derivatives of the vector $P(\theta)$. Furthermore, computation of $[\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}\dot{P}(\theta_n^*)]^{-1}$ would give an estimate of the limiting variance-covariance matrix of the B.A.N. estimate $\hat{\theta}_n$.

This method would be good for example in estimating the parameters of a Neyman type A distribution, where the vector $P(\theta)$ is a rather complicated function of the parameters, and other methods of getting B.A.N. estimates are rather difficult. This method has been applied by Robert Read of the Statistical Laboratory of the University of California, to estimating the parameters in a probabilistic model describing ionization in a cloud chamber, using as the preliminary estimates, those given by the method of moments. It has also been applied by Dr. Irene Rosenthal of the Psychology Department at the University of California, to estimate the parameters of a latent structure, using as first estimates those of Lazarsfeld [11].

4. Application to the problem of estimating bacterial density by the dilution method. The method of estimating the bacterial density of a liquid by taking samples in fermentation tubes at several levels of dilution of the liquid is well known. As far back as 1915 [12] the maximum likelihood estimate, called the

most probable number (M.P.N.) by Biometricians, was suggested for the problem, and is still being used today in Public Health for water, milk, and sewage analysis. This and other estimates have been studied by Fisher [13], Halvorson and Ziegler [14], and Matuszewski, Neyman, and Supinska [15].

The situation is the following. We are given a large volume V of a liquid containing a large number N of bacteria, and we are interested in estimating the bacterial density $\lambda = N/V$, the number of bacteria per unit volume. A sample of size α unit volume is withdrawn and tested by some device such as placing the sample in a fermentation tube to see if any bacteria are present. It is assumed that each bacterium acts independently and that each has the same probability α/V of being in the sample. Thus the number of bacteria in the sample will be binomially distributed with probability α/V and size N; however, if α/V is small and N is large the distribution may conveniently be replaced by a Poisson with parameter $N\alpha/V = \alpha\lambda$. The probability that no bacteria appear in the sample is then $p = e^{-\alpha\lambda}$. If n independent samples of size α are withdrawn and tested, the number K of sterile samples will be binomially distributed with probability p and size n, and may be used to estimate the parameter λ. However, the value of the experiment depends to a great extent on choosing α so that $p = e^{-\alpha\lambda}$ will be in a good estimating range, for if p is too small or too close to one, one will obtain too many fertile or too many sterile samples to be able to estimate λ with much accuracy. And since λ is unknown it will usually be impossible to choose α so that $e^{-\alpha\lambda}$ will be moderately between zero and one. So one usually takes several sizes of sample volumes α_1, $\alpha_2, \cdots, \alpha_s$, called dilution levels, and numbers of samples $n_1, n_2, \cdots, n_s$ at each of the levels, with the hope that at least one of the $e^{-\alpha_i\lambda}$ will be in a good estimating range. Then the numbers $k_1, k_2, \cdots, k_s$, of sterile samples at each of the levels will be used to estimate λ.

The most frequently used B.A.N. estimate of the bacterial density seems to be the maximum likelihood estimate, since the minimum χ^2 estimates appear to be much more difficult to compute. The maximum likelihood estimate of λ is that value of λ which is a root of the equation

$$\sum_{i=1}^{s} \frac{(n_i - k_i)\alpha_i}{(1 - e^{-\alpha_i\lambda})} = \sum_{i=1}^{s} n_i\alpha_i. \tag{4.1}$$

Methods of solving this equation have been discussed by Halvorson and Zeigler [14], Barkworth and Irwin [16], and Finney [17]. Tables of the estimate for certain situations may be found in Halvorson and Zeigler and in Hoskins [18].

An application of the methods of the previous section will yield a B.A.N. estimate which is slightly easier to compute. Linear forms which lead to B.A.N. estimates are of the type

$$\sum_{i=1}^{s} n_i f_i(z, \lambda)(z_i - e^{-\alpha_i\lambda}) \tag{4.2}$$

where z_i represents the frequency of sterile tubes at the ith level of dilution, $z_i = k_i/n_i$, and $f_i(z, \lambda)$ converges in probability to $\alpha_i(1 - e^{-\alpha_i\lambda})^{-1}$, z representing the vector $(z_1, \cdots, z_s)$. Equation (4.2) with $f_i(z, \lambda)$ always equal to

$$\alpha_i(1 - e^{-\alpha_i\lambda})^{-1}$$

is equivalent to the maximum likelihood equation (4.1).

We would like to replace $f_i(z, \lambda)$ in equation (4.2) completely by an estimate, that is, $f_i(z, \lambda) = \alpha_i/(1 - z_i)$, but we must take care of the cases in which z_i is equal to one. So we may choose $f_i(z, \lambda) = \alpha_i/(1 - z_i)$ if $z_i \neq 1$ and

$$f_i(z, \lambda) = \alpha_i(1 - e^{-\alpha_i\lambda})^{-1}$$

if $z_i = 1$. This will lead to a B A.N estimate since eventually as the n_i get large without bound, all the z_i will be different from one. We have the equation

$$\sum_{z_i \neq 1} n_i \frac{\alpha_i}{1 - z_i}(z_i - e^{-\alpha_i\lambda}) + \sum_{z_i = 1} n_i\alpha_i = 0. \tag{4.3}$$

Written in simpler form, this equation becomes

$$\sum_{z_i \neq 1} n_i \frac{\alpha_i}{1 - z_i} e^{-\alpha_i\lambda} = \sum_{z_i \neq 1} n_i \frac{\alpha_i}{1 - z_i} z_i + \sum_{z_i = 1} n_i\alpha_i. \tag{4.4}$$

This equation is simpler to solve than equation (4 1) in that it only requires tables of e^{-x} which are readily available, while equation (4.1) requires for its solution the computation of $(1 - e^{-\alpha_i\lambda})^{-1}$ separately for each i or tables of $(e^x - 1)^{-1}$ or $(1 - e^{-x})^{-1}$. The method by which it is suggested that (4.4) be solved is the same as that suggested by other authors in connection with the solution of (4.1), and that is Newton's method. For a function $f(x)$ with a continuous first derivative, if x_0 is taken to be the initial guess at the solution of $f(x) = 0$, x_n is defined inductively by

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}. \tag{4.5}$$

Applying this procedure to equation (4.4), we obtain the inductive formula

$$\lambda_n = \lambda_{n-1} + \frac{\sum_{z_i \neq 1} \frac{n_i\alpha_i}{1 - z_i} e^{-\alpha_i\lambda_{n-1}} - \sum_{z_i \neq 1} \frac{n_i\alpha_i}{1 - z_i} z_i - \sum_{z_i = 1} n_i\alpha_i}{\sum_{z_i \neq 1} \frac{n_i\alpha_i^2}{1 - z_i} e^{-\alpha_i\lambda_{n-1}}}. \tag{4.6}$$

The author has made a numerical study of the small sample properties of this estimate, the minimum χ^2 estimate and the maximum likelihood estimate, which he hopes to publish at a later date. An indication is given in this study that in general the estimate given by equation (4.4) has slightly better small sample properties in the sense of bias and root mean square error, than either the maximum likelihood or the minimum χ^2 estimate.

In conclusion, I would like to express my thanks to Professor L. Le Cam for his generous advice and helpful discussions concerning this paper.

REFERENCES

[1] J. NEYMAN, "Contribution to the theory of the χ^2 test," *Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press (1949).

[2] E. BARANKIN AND J. GURLAND, "On asymptotically normal efficient estimators: I," *University of California Publications in Statistics*, Vol. 1, No. 6 (1951), pp. 86-130.

[3] C. L. CHIANG, "On regular best asymptotically normal estimates with an application to a stochastic process," Doctoral Dissertation, University of California (1952).

[4] L. LE CAM, Lecture notes on large sample theory (mimeographed), University of California, 1955.

[5] W. TAYLOR, "Distance functions and regular best asymptotically normal estimates," *Ann. Math. Stat.*, Vol. 24, No. 1 (1953), pp. 85–92.

[6] J. BERKSON, "Application of the logistic function to bio-assay," *J. Amer. Stat. Assn.*, Vol. 39 (1944), pp. 357–365.

[7] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press (1946).

[8] J. PIERPONT, *The Theory of Functions of Real Variables*, Vol. I, Ginn and Co., (1905).

[9] L. M. GRAVES, *The Theory of Functions of Real Variables*, McGraw-Hill Book Co., Inc. (1946).

[10] H. CHERNOFF, "Large sample theory: parametric case," *Ann. Math. Stat.*, Vol. 27, No. 1 (1956), pp. 1–22.

[11] S. A. STOUFFER, *Measurement and Prediction*, Princeton University Press (1950).

[12] M. H. MCCRADY, "The numerical interpolation of fermentation-tube results," *Jour. Infec. Dis.*, Vol. 17 (1915), pp. 183–212.

[13] R. A. FISHER, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Roy. Soc., London, Ser. A.*, Vol. 222 (1922), pp. 309–368.

[14] H. O. HALVORSEN AND N. R. ZIEGLER, "Application of statistics to problems in bacteriology," *Jour. Bact.*, Vol. 25 (1933), pp. 101–121.

[15] T. MATUSZEWSKI, J. NEYMAN, AND J. SUPINSKA, "Statistical studies in questions of bacteriology," *J. Roy. Stat. Soc., Ser. B.*, Vol. 2 (1932), pp. 63–82.

[16] H. BARKSWORTH AND J. O. IRWIN, "Distribution of coliform organisms in milk and the accuracy of the presumptive coliform test," *Jour. Hyg.*, Cambridge, Vol. 38 (1938), pp. 446–457.

[17] D. J. FINNEY, "The principle of biological assay," *J. Roy. Stat. Soc., Ser. B.*, Vol. 9 (1947), pp. 46–91.

[18] J. K. HOSKINS, "Most probable numbers for evaluation of coli aerogenes tests by fermentation tube method," Public Health Reports, Vol. 49 (1934), p. 393, Reprint No. 1621 (revised 1940).

FAMILIES OF DESIGNS FOR TWO SUCCESSIVE EXPERIMENTS

BY G. H. FREEMAN

East Malling Research Station, Maidstone, England

It is sometimes desirable, particularly in experimentation with perennial plants, to conduct an experiment on plots already used for a previous trial. Various designs are known that facilitate this process (Hoblyn *et al.*, 1954), the following notation being used to describe types of design. The letters O, T and P refer respectively to designs that are orthogonal, totally balanced—i e., balanced incomplete blocks—and partially balanced; then a design of type X:YZ where X, Y or Z may be any of O, T or P, is one in which the arrangement of the first set of treatments with respect to blocks is of type X, that of the second set of treatments to blocks is of type Y and that of the second set of treatments to the first is of type Z. It is assumed that the two sets of treatments are non-interacting and that designs of type T or P may be extended, i e , have complete replicates in each block. Then, if the first trial is in randomised blocks, type O, the only design that has not previously been very fully discussed is type O:PP for which, however, general methods of analysis have been given (Freeman, 1957b). The purpose of this paper is to describe all known families of O:PP designs with two associate-classes. These, being designs with two orthogonal constraints, can also be regarded as row and column designs, and will henceforth be considered as such here.

The families of O:PP designs described here include all those with any members likely to be of much practical use, i.e., having more than two replicates or treatments, not more than 30 replicates, treatments, rows or columns and not more than 150 plots in all. The possibilities of existence of all O:PP designs within these limits have been investigated by enumeration and all the tabulated designs have been found to exist. Where larger designs are required their existence can usually be readily determined and, particularly where the number of replicates greatly exceeds the number of treatments, there may be many possible designs. A catalogue of the designs in Tables II–V and VII has been prepared, and is available at East Malling Research Station; the construction of an individual design gives rise to no practical difficulty by trial and error, but no attempt has been made to find transformation sets for each design.

Since, for practical purposes, O:PP designs are constrained to have the same associate-classes with respect to rows and columns their classification depends on that of designs of type P. These, partially balanced, designs have been described in great detail by Bose and his co-workers, who have provided an extensive catalogue of such designs with two associate-classes, (Bose *et al.*, 1954). Although this catalogue is now known not to be exhaustive (see, for example, Archbold and Johnson, 1956 and Freeman, 1957a) it does provide a basis for the

Received June 7, 1957.

classification of designs of type P and is thus adopted here for O:PP designs as well.

Group Divisible Designs

The simplest O:PP designs are those that are group divisible, the property of group divisibility being such that designs group divisible one way are also group divisible the other. Group divisible partially balanced designs can be divided into the three types, singular, semi-regular and regular, of which the first and last will be denoted as S[1] and R respectively, as usual, while semi-regular designs will be called H, so as to keep a one-letter code. As each type of group divisible design can be combined with itself or each other to make an O:PP design there are thus six types of group divisible O:PP design, these being described, in an obvious notation, as SS, HH, HS, RR, RH, RS, and considered in this order.

To classify group divisible O:PP designs into families it is first necessary to consider the various types of singular, semi-regular and regular designs.

Singular designs. Families of singular designs are uniquely determined by balanced incomplete block designs (Bose and Connor, 1952), and thus a complete classification of the former is afforded by a corresponding one of the latter. The following types of balanced incomplete block design are considered, these not including every possible design but containing all that give rise to singular designs from which can be constructed O:PP designs of practicable size. We shall use the notation $C[n; k]$ for the binomial coefficient $\binom{n}{k}$

(i) $C[v - 1; k - 1]$ replicates of v treatments on $C[v; k]$ blocks of k plots each (unreduced)
(ii) $(s + 1)$ replicates of s^2 treatments on $s(s + 1)$ blocks of s plots each (orthogonal series 1 or OS 1)
(iii) $(s^2 - 1)$ replicates of s^2 treatments on $s(s + 1)$ blocks of $s(s - 1)$ plots each (complement of OS 1)
(iv) $(2t + 1)$ replicates of $(2t + 2)$ treatments on $(4t + 2)$ blocks of $(t + 1)$ plots each
(v) $(2t + 1)$ replicates of $(4t + 3)$ treatments on $(4t + 3)$ blocks of $(2t + 1)$ plots each
(vi) $(2t + 2)$ replicates of $(4t + 3)$ treatments on $(4t + 3)$ blocks of $(2t + 2)$ plots each.

The first of these types exists for all values of v and all $k < v$, types (ii) and (iii) for all values of s for which complete sets of orthogonal Latin squares exist, and types (iv)–(vi) when $(4t + 3)$ is a prime-power (Bose, 1939); Bose also shows geometrically that, though $(4t + 3)$ is not a prime-power for $t = 3$, designs of

[1] The letters S here and T for triangular designs used later in the paper are also used respectively for designs with supplemented balance and total balance by Hoblyn *et al.* (1954), but there should be no confusion between the uses for *types* of balanced designs and *families* of partially balanced designs.

types (iv)–(vi) are possible for $t = 3$. Types (v) and (vi) include the second orthogonal series (OS 2) and its complement, respectively, when $t = 1$, the corresponding value of s in the orthogonal series designs being $s = 2$.

The classification of balanced incomplete block designs into these six types is not mutually exclusive; for example, the design with 3 replicates of 4 treatments on 6 blocks of 2 plots each can be considered as belonging to each of types (i), (ii), (iii) and (iv). It follows that particular O:PP designs singular one way may belong to more than one family. For the sake of uniqueness, any one design of type P or O:PP will be regarded as a member of only one family, since the overlapping of families is a direct consequence of the overlap of types of balanced incomplete block design and thus irrelevant to the consideration of O:PP designs.

A balanced incomplete block design with r^* replicates of v^* treatments on b^* blocks of k^* plots each gives an unextended singular P design with r^* replicates of nv^* treatments on b^* blocks with nk^* plots each, there being v^* groups of n treatments each, $n > 1$. Thus, on allowing extended designs with p complete replicates of treatments in each block and, further, the whole design repeated q times, the most general singular P has the following numbers of replicates, treatments, blocks and plots per block: $q(r^* + pb^*)$, nv^*, qb^*, $n(k^* + pv^*)$. The parameters of all designs of type P will be written in this order henceforth; m will be used instead of v^* so as to consider m groups of treatments in conformity with the usual notation.

The types of balanced incomplete block design enumerated above give rise to the following families of singular designs, where the inequalities are inserted, in S(ii)–(vi), to ensure the uniqueness of the families:

S(i) $\quad \dfrac{q(pm + k)}{m} C[m, k], mn, qC[m;k], n(pm + k) \qquad (k < m)$

S(ii) $\quad q(s + 1)(ps + 1), ns^2, qs(s + 1), ns(ps + 1) \qquad (s > 2)$

S(iii) $\quad q(s + 1) [s(p + 1) - 1], ns^2, qs(s + 1), ns[s(p + 1) - 1] \qquad (s > 2)$

S(iv) $\quad q(2t + 1)(2p + 1), 2n(t + 1), 2q(2t + 1), n(t + 1)(2p + 1) \qquad (t > 1)$

S(v) $\quad q[2t(2p + 1) + 3p + 1], n(4t + 3), q(4t + 3), n[2t(2p + 1) + 3p + 1] \qquad (t > 0)$

S(vi) $\quad q[2t(2p + 1) + 3p + 2], n(4t + 3), q(4t + 3), n[2t(2p + 1) + 3p + 2] \qquad (t > 0)$

Semi-regular designs. Bose *et al.* (1953) classify semi-regular designs according as λ_1 does or does not equal zero, but this classification seems unnecessary for the present purpose, $\lambda_1 = 0$ being merely a special case. Thus, from Bose *et al.*, the design has $n\lambda_2/c$ replicates of mn treatments in m groups of n on $n^2\lambda_2/c^2$ blocks of mc plots, where $\lambda_1 = n\lambda_2(c - 1)/c(n - 1)$ is integral and $m \leq (n^2\lambda_2 - c^2)/c^2(n - 1)$. The extension of this design to allow for p complete replicates of the

treatments in each block and the whole design repeated q times gives the following family of semi-regular designs:

$$\text{H} \qquad \frac{qn\lambda_2(pn + c)}{c^2}, mn, \frac{qn^2\lambda_2}{c^2}, m(pn + c), \text{ where } n\frac{\lambda_2(c - 1)}{c(n - 1)}$$

is integral and

$$m \leqq \frac{n^2\lambda_2 - c^2}{c^2(n - 1)}$$

Regular designs. These are more difficult to categorise than either of the other types of group divisible design. Further, practicable O:PP designs cannot be derived from all types of regular design; two such types are those generated by the methods of differences and of omitting varieties (Bose *et al.*, 1953), and thus these two types are not considered further.

The types of regular design that are considered are as follows:

(i) designs derivable by addition,

(ii) designs with complete and incomplete groups,

(iii) designs with groups arranged in sets,

(iv) disconnected designs. The first of these consists of all designs derivable by addition of group divisible designs to other group divisible designs or to balanced incomplete block designs, while the next two types are described elsewhere (Freeman, 1957a). The fourth type is not considered in an unextended form, as it has been shown (Freeman, 1957c) that this type of design cannot give rise to an O:PP design; however, extended disconnected designs are of use for the construction of O:PP designs.

In order to consider only those designs of type P that give rise to O:PP designs when the plots within each block are rearranged in accordance with a second classification further restrictions on the parameters are necessary in types R(ii) and R(iv). With these further restrictions, the four families of regular designs are as follows, there being p complete replicates of the treatments in each block and the whole design being repeated q times:

$$\text{R(i)} \qquad \frac{qR(k + pmn)}{k}, mn, \frac{qRmn}{k}, k + pmn,$$

where

$R = ar + a'r'$, $a, r, a', r' > 0$, $k > 1$,

$$\lambda_1(n - 1) + \lambda_2 n(m - 1) = r(k - 1),\ \lambda_1'(n - 1) + \lambda_2' n(m - 1) = r'(k - 1),$$

$$a\lambda_1 + a'\lambda_1' = \Lambda_1 \neq a\lambda_2 + a'\lambda_2' = \Lambda_2.$$

$$\text{R(ii)} \qquad \frac{qC[m - 1; u]C[n; h](nu + h + pmn)}{n}, mn,$$

$$qmC[m - 1; u]C[n; h], nu + h + pmn,$$

where $0 < u < m$, $1 < h < n - 1$.

R(iii) $$q(3n - 1)(pn + 1),\ 2n^2,\ qn(3n - 1),\ 2n(pn + 1).$$

R(iv) $$\frac{qr(k + pmn)}{k},\ mn,\ \frac{qrmn}{k},\ k + pmn,$$

where $r(k - 1)/(n - 1)$ is integral, $p > 0$, $1 < k < n - 1$.

There are m groups of n treatments in each design, where, for family R(iii), $m = 2n$.

Construction of O:PP designs. Not every pair of designs of type P can give rise to an O:PP design. Thus, families S(ii) and S(iii) are incompatible with S(v) and S(vi), as s^2 is never of the form $(4t + 3)$. Also R(iii) is incompatible with R(ii) and R(iv) as, in order to satisfy the relations $qn(3n - 1) = nu + h + pmn$ or $qn(3n - 1) = k + pmn$, h or k must be a multiple of n, an impossibility since each lies between 1 and $(n - 1)$. Further, S(iv) and R(iii) require an even number of groups of treatments and so cannot be associated with S(v) and S(vi), which require an odd number.

All the families of group divisible O:PP designs not excluded by the above argument are given in Table I together with their derivation from the corresponding families of P designs. As an example, consider family SS II, derived from families S(ii) and S(i). For the numbers of groups of treatments to be the same in the two designs the relation $m = s^2$ must be satisfied, while for the blocks and plots per block of the two families to be interchangeable two further relations must hold. Throughout Table I, to distinguish between the parameters in the two families of type P, that written second has dashes, and so here p' and q' refer to S(i) while p and q refer to S(ii). The relations between blocks and plots per block then are: $qs(s + 1) = n(p's^2 + k)$, $q'C[s^2; k] = ns(ps + 1)$. Thus

$$q = \frac{n(p's^2 + k)}{s(s + 1)} \text{ and } q' = \frac{ns(ps + 1)}{C[s^2; k]}$$

as shown in Table I. The number of treatments in the design is ns^2, while the last two columns of Table I show the numbers of blocks in the two designs of type P, i.e., $qs(s + 1)$ for S(ii) and $q'C[s^2; k]$ for S(i). The number of replicates is shown in Table I as $q(s + 1)(ps + 1)$, that corresponding to S(ii), although it could be given in several forms; by convention, the number is given throughout Table I in terms of the replicates of the family of type P given first.

In Table I, all the numbers shown are non-negative integers and are subject to the restrictions described in the classification above of designs of type P. In certain families further restrictions are necessary on the values of the parameters by virtue of the first of the non-existence theorems (Freeman, 1957c); thus in any family constructed from S(i) the number of treatments in the O:PP design must be greater than or equal to the number of rows plus the number of columns, while in family SS III p and p' cannot both be zero. In families SS VIII and SS IX, in order to satisfy the relation $s^2 = 2(t + 1)$, the parameter w is intro-

TABLE I

Families of Group Divisible O:PP Designs

Family	Derivation (a) x (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
SS I	S(i) x S(i)	$\frac{n(p'm + k')}{C[m;k]}$	$\frac{n(pm + k)}{C[m;k']}$	$\frac{q(pm + k)}{m} C[m;k]$	mn	$qC[m;k]$	$q'C[m;k']$
SS II	S(ii) x S(i)	$\frac{n(p's^2 + k)}{s(s + 1)}$	$\frac{ns(ps + 1)}{C[s^2;k]}$	$q(s + 1)(ps + 1)$	ns^2	$qs(s + 1)$	$q'C[s^2;k]$
SS III	S(ii) x S(ii)	$\frac{n(p's + 1)}{s + 1}$	$\frac{n(ps + 1)}{s + 1}$	$q(s + 1)(ps + 1)$	ns^2	$qs(s + 1)$	$q's(s + 1)$
SS IV	S(iii) x S(i)	$\frac{n(p's^2 + k)}{s(s + 1)}$	$\frac{ns[s(p + 1) - 1]}{C[s^2;k]}$	$q(s + 1)[s(p + 1) - 1]$	ns^2	$qs(s + 1)$	$q'C[s^2;k]$
SS V	S(iii) x S(ii)	$\frac{n(p's + 1)}{s + 1}$	$\frac{n[s(p + 1) - 1]}{s + 1}$	$q(s + 1)[s(p + 1) - 1]$	ns^2	$qs(s + 1)$	$q's(s + 1)$
SS VI	S(iii) x S(iii)	$\frac{n[s(p' + 1) - 1]}{s + 1}$	$\frac{n[s(p + 1) - 1]}{s + 1}$	$q(s + 1)[s(p + 1) - 1]$	ns^2	$qs(s + 1)$	$q's(s + 1)$
SS VII	S(iv) x S(i)	$\frac{n[2p'(t + 1) + k]}{2(2t + 1)}$	$\frac{n(t + 1)(2p + 1)}{C[2(t + 1);k]}$	$q(2t + 1)(2p + 1)$	$2n(t + 1)$	$2q(2t + 1)$	$q'C[2(t + 1);k]$
SS VIII	S(iv) x S(ii)	$\frac{nw(2p'w + 1)}{4w^2 - 1}$	$\frac{nw(2p + 1)}{2w + 1}$	$q(4w^2 - 1)(2p + 1)$	$4nw^2$	$2q(4w^2 - 1)$	$2q'w(2w + 1)$
SS IX	S(iv) x S(iii)	$\frac{nw[2w(p' + 1) - 1]}{4w^2 - 1}$	$\frac{nw(2p + 1)}{2w + 1}$	$q(4w^2 - 1)(2p + 1)$	$4nw^2$	$2q(4w^2 - 1)$	$2q'w(2w + 1)$
SS X	S(iv) x S(iv)	$\frac{n(t + 1)(2p' + 1)}{2(2t + 1)}$	$\frac{n(t + 1)(2p + 1)}{2(2t + 1)}$	$q(2t + 1)(2p + 1)$	$2n(t + 1)$	$2q(2t + 1)$	$2q'(2t + 1)$
SS XI	S(v) x S(i)	$\frac{n[p'(4t + 3) + k]}{4t + 3}$	$\frac{n[2t(2p + 1) + 3p + 1]}{C[4t + 3;k]}$	$q[2t(2p + 1) + 3p + 1]$	$n(4t + 3)$	$q(4t + 3)$	$q'C[4t + 3;k]$
SS XII	S(v) x S(v)	$\frac{n[2t(2p' + 1) + 3p' + 1]}{4t + 3}$	$\frac{n[2t(2p + 1) + 3p + 1]}{4t + 3}$	$q[2t(2p + 1) + 3p + 1]$	$n(4t + 3)$	$q(4t + 3)$	$q'(4t + 3)$

TABLE I (Continued)

Design	Derivation (a) x (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
[illegible]	[illegible]	$\frac{n[p'(4t+3)+k]}{4t+3}$	$\frac{n[2t(2p+1)+3p+2]}{C[4t+3,k]}$	$q[2t(2p+1)+3p+2]$	$n(4t+3)$	$q(4t+3)$	$q'C[4t+3,k]$
[illegible]	[illegible]	[illegible]	$\frac{n[2t(2p+1)+3p+2]}{4t+3}$	$q[2t(2p+1)+3p+2]$	$n(4t+3)$	$q(4t+3)$	$q'(4t+3)$
[illegible]	[illegible]	[illegible]	$\frac{n[2t(2p+1)+3p+2]}{4t+3}$	$q[2t(2p+1)+3p+2]$	$n(4t+3)$	$q(4t+3)$	$q'(4t+3)$
NN	N x N	[illegible]	[illegible]	$\frac{qn\lambda_2(pn+c)}{c^2}$	mn	$\frac{qn^2\lambda_2}{c^2}$	$\frac{q'n^2\lambda_2'}{c'^2}$
NS I	N x S(i)	[illegible]	[illegible]	$\frac{qn\lambda_2(pn+c)}{c^2}$	mn	$\frac{qn^2\lambda_2}{c^2}$	$q'C[m,k]$
NS II	N x S(ii)	[illegible]	[illegible]	$\frac{qn\lambda_2(pn+c)}{c^2}$	ns^2	$\frac{qn^2\lambda_2}{c^2}$	$q's(s+1)$
NS III	N x S(iii)	[illegible]	[illegible]	[illegible]	ns^2	$\frac{qn^2\lambda_2}{c^2}$	$q's(s+1)$
NS IV	N x S(iv)	[illegible]	[illegible]	[illegible]	$2n(t+1)$	$\frac{qn^2\lambda_2}{c^2}$	$2q'(2t+1)$
NS V	N x S(v)	[illegible]	[illegible]	[illegible]	$n(4t+3)$	$\frac{qn^2\lambda_2}{c^2}$	$q'(4t+3)$
NS VI	N x S(vi)	[illegible]	[illegible]	[illegible]	[illegible]	$\frac{qn^2\lambda_2}{c^2}$	$q'(4t+3)$

TABLE I (Continued)

Family	Derivation (a) × (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
FF I	R(i) × R(i)	$\frac{k(k' + p'mn)}{Rmn}$	$\frac{k'(k + pmn)}{R'mn}$	$\frac{qR(k + pmn)}{k}$	mn	$\frac{qRmn}{k}$	$\frac{q'R'mn}{k'}$
FF II	R(ii) × R(i)	$\frac{k + p'mn}{mC[m-1;u]C[n;h]}$	$\frac{k(nu + h + pmn)}{Rmn}$	$\frac{qC[m-1;u]C[n;h](nu + h + pmn)}{n}$	mn	qmC[m−1;u]C[n;h]	$\frac{q'Rmn}{k}$
FF III	R(ii) × R(ii)	$\frac{nu' + h' + p'mn}{mC[m-1,u]C[n,h]}$	$\frac{nu + h + pmn}{mC[m-1;u']C[n;h']}$	$\frac{qC[m-1;u]C[n;h](nu + h + pmn)}{n}$	mn	qmC[m−1;u]C[n;h]	q'mC[m−1;u']C[n;h']
FF IV	I(iii) × R(i)	$\frac{k + 2p'n^2}{n(3n - 1)}$	$\frac{k(pn + 1)}{Rn}$	q(3n − 1)(pn + 1)	$2n^2$	qn(3n − 1)	$\frac{2q'Rn^2}{k}$
FF V	I(iii) × R(iii)	$\frac{2(p'n + 1)}{3n - 1}$	$\frac{2(pn + 1)}{3n - 1}$	q(3n − 1)(pn + 1)	$2n^2$	qn(3n − 1)	q'n(3n − 1)
FF VI	R(iv) × R(i)	$\frac{k(k' + p'mn)}{rmn}$	$\frac{k'(k + pmn)}{Rmn}$	$\frac{qr(k + pmn)}{k}$	mn	$\frac{qrmn}{k}$	$\frac{q'Rmn}{k'}$
FF VII	I(iv) × I(ii)	$\frac{k(nu + h + p'mn)}{rmn}$	$\frac{k + pmn}{mC[m-1;u]C[n,h]}$	$\frac{qr(k + pmn)}{k}$	mn	$\frac{qrmn}{k}$	q'mC[m−1;u]C[n;h]
FF VIII	I(iv) × I(iv)	$\frac{k(k' + p'mn)}{rmn}$	$\frac{k'(k + pmn)}{r'mn}$	$\frac{qr(k + pmn)}{k}$	mn	$\frac{qrmn}{k}$	$\frac{q'r'mn}{k'}$
[illegible]	I(i) × B	$\frac{k(p'n + c)}{Rn}$	$\frac{c^2(k + pmn)}{n^2\lambda_2}$	$\frac{qR(k + pmn)}{k}$	mn	$\frac{qRmn}{k}$	$\frac{q'n^2\lambda_2}{c^2}$
[illegible]	I(ii) × B	$\frac{p'n + c}{C[m-1,u]C[n,h]}$	$\frac{c^2(nu + h + pmn)}{n^2\lambda_2}$	$\frac{qC[m-1;u]C[n;h](nu + h + pmn)}{n}$	mn	qmC[m−1;u]C[n;h]	$\frac{q'n^2\lambda_2}{c^2}$
[illegible]	I(iii) × B	$\frac{2(p'n + c)}{3n - 1}$	$\frac{2c^2(pn + 1)}{n\lambda_2}$	q(3n − 1)(pn + 1)	$2n^2$	qn(3n − 1)	$\frac{q'n^2\lambda_2}{c^2}$
[illegible]	[illegible]	[illegible]	[illegible]	$\frac{qr(k + pmn)}{k}$	mn	$\frac{qrmn}{k}$	$\frac{q'n^2\lambda_2}{c^2}$

TABLE I (Continued)

Family	Derivation (a) x (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
RS I	R(i) x S(i)	$\frac{k(p'm + k')}{Rm}$	$\frac{k + pmn}{C[m,k']}$	$\frac{qR(k + pmn)}{k}$	mn	$\frac{qRmn}{k}$	$q'C[m,k']$
RS II	R(i) x S(ii)	$\frac{k(p's + 1)}{Rs}$	$\frac{k + pns^2}{s(s + 1)}$	$\frac{qR(k + pns^2)}{k}$	ns^2	$\frac{qRns^2}{k}$	$q's(s + 1)$
RS III	R(i) x S(iii)	$\frac{k[s(p' + 1) - 1]}{Rs}$	$\frac{k + pns^2}{s(s + 1)}$	$\frac{qR(k + pns^2)}{k}$	ns^2	$\frac{qRns^2}{k}$	$q's(s + 1)$
RS IV	R(i) x S(iv)	$\frac{k(2p' + 1)}{2R}$	$\frac{k + 2pn(t + 1)}{2(2t + 1)}$	$\frac{qR[k + 2pn(t + 1)]}{k}$	$2n(t + 1)$	$\frac{2qRn(t + 1)}{k}$	$2q'(2t + 1)$
RS V	R(i) x S(v)	$\frac{k[2t(2p' + 1) + 3p' + 1]}{R(4t + 3)}$	$\frac{k + pn(4t + 3)}{4t + 3}$	$\frac{qR[k + pn(4t + 3)]}{k}$	$n(4t + 3)$	$\frac{qRn(4t + 3)}{k}$	$q'(4t + 3)$
RS VI	R(i) x S(vi)	$\frac{k[2t(2p' + 1) + 3p' + 2]}{R(4t + 3)}$	$\frac{k + pn(4t + 3)}{4t + 3}$	$\frac{qR[k + pn(4t + 3)]}{k}$	$n(4t + 3)$	$\frac{qRn(4t + 3)}{k}$	$q'(4t + 3)$
RS VII	R(ii) x S(i)	$\frac{n(p'm + k)}{mC[m-1,u]C[n,h]}$	$\frac{nu + h + pmn}{C[m,k]}$	$\frac{qC[m-1,u]C[n,h](nu + h + pmn)}{n}$	mn	$qmC[m-1,u]C[n,h]$	$q'C[m,k]$
RS VIII	R(ii) x S(ii)	$\frac{n(p's + 1)}{sC[s^2-1,u]C[n,h]}$	$\frac{nu + h + pns^2}{s(s + 1)}$	$\frac{qC[s^2-1,u]C[n,h](nu + h + pns^2)}{n}$	ns^2	$qs^2C[s^2-1,u]C[n,h]$	$q's(s + 1)$
RS IX	R(ii) x S(iii)	$\frac{n[s(p' + 1) - 1]}{sC[s^2-1,u]C[n,h]}$	$\frac{nu + h + pns^2}{s(s + 1)}$	$\frac{qC[s^2-1,u]C[n,h](nu + h + pns^2)}{n}$	ns^2	$qs^2C[s^2-1,u]C[n,h]$	$q's(s + 1)$
RS X	R(ii) x S(iv)	$\frac{n(2p' + 1)}{2C[2t + 1,u]C[n,h]}$	$\frac{nu + h + 2pn(t + 1)}{2(2t + 1)}$	$\frac{qC[2t+1,u]C[n,h][nu+h+2pn(t+1)]}{n}$	$2n(t+1)$	$2q(t+1)C[2t+1,u]C[n,h]$	$2q'(2t + 1)$
RS XI	R(ii) x S(v)	$\frac{n[2t(2p' + 1) + 3p' + 1]}{(4t + 3)C[2(2t + 1),u]C[n,h]}$	$\frac{nu + h + pn(4t + 3)}{4t + 3}$	$\frac{qC[4t+2,u]C[n,h][nu+h+pn(4t+3)]}{n}$	$n(4t+3)$	$q(4t+3)C[4t+2,u]C[n,h]$	$q'(4t + 3)$

TABLE I (Continued)

Family	Derivation (a) x (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
ES XII	R(ii) x S(vi)	$\frac{n[2t(2p'+1)+3p'+2]}{(4t+3)C[2(2t+1);u]C[n;h]}$	$\frac{nu+h+pn(4t+3)}{4t+3}$	$\frac{qC[4t+2;u]C[n;h][nu+h+pn(4t+3)]}{n}$	$n(4t+3)$	$q(4t+3)C[4t+2;u]C[n;h]$	$q'(4t+3)$
ES XIII	R(iii) x S(i)	$\frac{2p'n+k}{3n-1}$	$\frac{2n(pn+1)}{C[2n;k]}$	$q(3n-1)(pn+1)$	$2n^2$	$qn(3n-1)$	$q'C[2n;k]$
ES XIV	R(iii) x S(ii)	$\frac{2w(2p'w+1)}{6w^2-1}$	$\frac{2w(2pw^2+1)}{2w+1}$	$q(6w^2-1)(2pw^2+1)$	$8w^4$	$2qw^2(6w^2-1)$	$2q'w(2w+1)$
ES XV	R(iii) x S(iii)	$\frac{2w[2w(p'+1)-1]}{6w^2-1}$	$\frac{2w(2pw^2+1)}{2w+1}$	$q(6w^2-1)(2pw^2+1)$	$8w^4$	$2qw^2(6w^2-1)$	$2q'w(2w+1)$
ES XVI	R(iii) x S(iv)	$\frac{(t+1)(2p'+1)}{3t+2}$	$\frac{(t+1)[p(t+1)+1]}{2t+1}$	$q(3t+2)[p(t+1)+1]$	$2(t+1)^2$	$q(t+1)(3t+2)$	$2q'(2t+1)$
ES XVII	F(iv) x S(i)	$\frac{k(p'n+k')}{rm}$	$\frac{k+pmn}{C[m;k']}$	$\frac{qr(k+pmn)}{k}$	mn	$\frac{qrmn}{k}$	$q'C[m;k']$
ES XVIII	R(iv) x S(ii)	$\frac{k(p's+1)}{rs}$	$\frac{k+pns^2}{s(s+1)}$	$\frac{qr(k+pns^2)}{k}$	ns^2	$\frac{qrns^2}{k}$	$q's(s+1)$
ES XIX	R(iv) x S(iii)	$\frac{k[s(p'+1)-1]}{rs}$	$\frac{k+pns^2}{s(s+1)}$	$\frac{qr(k+pns^2)}{k}$	ns^2	$\frac{qrns^2}{k}$	$q's(s+1)$
ES XX	R(iv) x S(iv)	$\frac{k(2p'+1)}{2r}$	$\frac{k+2pn(t+1)}{2(2t+1)}$	$\frac{qr[k+2pn(t+1)]}{k}$	$2n(t+1)$	$\frac{2qrn(t+1)}{k}$	$2q'(2t+1)$
ES XXI	R(iv) x S(v)	$\frac{k[2t(2p'+1)+3p'+1]}{r(4t+3)}$	$\frac{k+pn(4t+3)}{4t+3}$	$\frac{qr[k+pn(4t+3)]}{k}$	$n(4t+3)$	$\frac{qrn(4t+3)}{k}$	$q'(4t+3)$
ES XXII	R(iv) x S(vi)	$\frac{k[2t(2p'+1)+3p'+2]}{r(4t+3)}$	$\frac{k+pn(4t+3)}{4t+3}$	$\frac{qr[k+pn(4t+3)]}{k}$	$n(4t+3)$	$\frac{qrn(4t+3)}{k}$	$q'(4t+3)$

duced, where $s = 2w$ and so $(t + 1) = 2w^2$; similarly, in RS XIV and RS XV to satisfy $s^2 = 2n$, w is again introduced, with $s = 2w$ and $n = 2w^2$. These uses of w are the only occasions where auxiliary symbols are necessary in this Table

Useful group divisible designs. Tables II–V give the numerical values of the parameters of all useful group divisible designs. Where there is a design with the same parameters but simpler than O:PP the O:PP design is not included in the appropriate Table. In these Tables, which are derived by putting numerical values in Table I, "rows" and "columns" are substituted for the two types of "blocks" of the basic designs of type P. Rows and columns are chosen not to correspond to either type of block but merely so that there are never more rows than columns, this convention being a help to the writing down of the designs themselves.

It will be seen that by far the greatest number of useful designs occurs in the three families SS I, HS I and RS I, the designs from which are enumerated in Tables II, III and IV respectively. As no other family has more than five useful designs, the useful designs in all other families are given in Table V, in which many of the parameters are inapplicable in each line.

Only ten families are represented in Table V, and so Table VI has been constructed from Table I showing the smallest designs that are theoretically possible in the other 43 families that have no useful designs. Since even the smallest of these, those for SS II and RR VI, have 216 plots, while the largest, that for RS XII, has 8744736 plots, no attempt has been made to discover whether these designs do in fact exist. None of them appears to be excluded by any of the non-existence theorems (Freeman, 1957c), and so they are all presumed to be possible; however none is likely to be practicable Table VI, like Table V, has many parameters inapplicable in each line.

In Tables IV, V and VI, whenever designs derived from the P design R(i) are given, the auxiliary parameters Λ_1 and Λ_2 are shown. This is because two designs, like RS I 1 and RS I 2 in Table IV, may differ from each other only in respect of these parameters. No other auxiliary parameters are necessary, as these are relevant only to the construction of the design, not its final form.

NON-GROUP DIVISIBLE DESIGNS

Like designs of type P, O:PP designs that are group divisible are much more numerous than those that are not. Of the other types of P design, only the triangular and Latin square appear to give rise to O:PP designs; while the other types may, none of the designs given by Bose *et al.* (1954) leads to an O:PP design and, at least for cyclic designs, it appears unlikely that any could. Since in an O:PP design the association scheme is the same both ways designs that are either triangular or Latin square one way are the same the other, the association schemes being unique for each type.

Triangular designs. The basic triangular design has r replicates of $n(n - 1)/2$ treatments on $rn(n - 1)/2k$ blocks of k plots each. If $n = 2$ the design has only

TABLE V

Other useful group divisible designs

Design	s	t	m	n	k	k′	c	λ_2	c′	λ_2'	R	Λ_1	Λ_2	R′	Λ_1'	Λ_2'	u	h	r	p	p′	q	q′	Rep	Tr	Rows	Cols
SS VI 1	3	-	9	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	1	8	18	12	12
SS VII 1	-	2	6	2	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	1	5	12	6	10
SS VII 2	-	3	8	2	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	1	7	16	8	14
HH 1	-	-	2	4	-	-	2	3	2	3	-	-	-	-	-	-	-	-	-	1	1	1	1	18	8	12	12
HH 2	-	-	4	2	-	-	1	3	1	3	-	-	-	-	-	-	-	-	-	1	1	1	1	18	8	12	12
HS V 1	-	1	7	2	-	-	1	5	-	-	-	-	-	-	-	-	-	-	-	0	1	1	1	10	14	7	20
HS VI 1	-	1	7	2	-	-	1	2	-	-	-	-	-	-	-	-	-	-	-	0	0	1	1	4	14	7	8
HS VI 2	-	2	11	2	-	-	1	3	-	-	-	-	-	-	-	-	-	-	-	0	0	1	1	6	22	11	12
RR I 1	-	-	2	2	2	2	-	-	-	-	5	1	2	5	1	2	-	-	-	2	2	1	1	25	4	10	10
RR I 2	-	-	2	2	2	2	-	-	-	-	5	1	2	5	3	1	-	-	-	2	2	1	1	25	4	10	10
RR I 3	-	-	2	2	2	2	-	-	-	-	5	3	1	5	3	1	-	-	-	2	2	1	1	25	4	10	10
RH I 1	-	-	2	3	3	-	1	1	-	-	7	4	2	-	-	-	-	-	-	1	2	1	1	21	6	9	14
RH I 2	-	-	2	3	3	-	2	4	-	-	8	5	2	-	-	-	-	-	-	1	2	1	1	24	6	9	16
RH I 3	-	-	3	2	4	-	1	1	-	-	14	10	8	-	-	-	-	-	-	0	3	1	1	14	6	4	21
RH I 4	-	-	3	2	4	-	1	1	-	-	18	10	11	-	-	-	-	-	-	0	4	1	1	18	6	4	27
RH I 5	-	-	3	2	4	-	1	1	-	-	18	14	10	-	-	-	-	-	-	0	4	1	1	18	6	4	27
RS VII 1	-	-	2	4	1	-	-	-	-	-	-	-	-	-	-	-	1	2	-	0	1	1	3	9	8	6	12
RS XIII 1	-	-	4	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1	1	5	8	4	10
RS XIII 2	-	-	4	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	3	15	8	10	12
RS XVII 1	-	-	2	4	2	1	-	-	-	-	-	-	-	-	-	-	-	-	3	1	1	1	5	15	8	10	12

one treatment, if $n = 3$ the design is one in balanced incomplete blocks and if $n = 4$ *it may be regarded as a group divisible design with 3 groups of 2 treatments* each; hence for practical purposes $n > 4$. Further, in addition to the possibilities of extension by adding p complete replicates to each block and repeating the whole design, a times say, complete replicates of balanced incomplete block designs may be added. Thus, in fully extended form, the basic triangular design is as follows:

$$\text{T} \qquad \frac{q[2k + pn(n-1)]}{2k}, \frac{n(n-1)}{2}, \frac{qn(n-1)}{2k}, \frac{2k + pn(n-1)}{2},$$

where $q = ar + AR$,

$$\lambda = \frac{2R(k-1)}{(n-1)(n-2)}, n > 4$$

Thus the triangular O:PP design has the following form.

$$\text{TT} \qquad \frac{q[2k + pn(n-1)}{2k}, \frac{n(n-1)}{2}, \frac{qn(n-1)}{2k}, \frac{q'n(n-1)}{2k'},$$

where

$$q = \frac{k[2k' + p'n(n-1)]}{n(n-1)}, n > 4,$$

$$q' = \frac{k'[2k + pn(n-1)]}{n(n-1)}$$

and there are the further restrictions above on q and similarly on q'

TABLE VI

Smallest possible designs in families with no useful members

Family	s	s'	t	m	n	k	k'	c	λ_1	R	Λ_1	Λ_2	a	b	a'	b'	r	r'	p	p'	q	q'	Rep.	Tr.	Rows	Cols.
SS II	3	-	-	9	3	3	-	-	-	-	-	-	-	-	-	-	-	-	0	0	2	1	8	27	9	24
SS III	3	-	-	9	2	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	2	2	32	18	21	24
SS IV	3	-	-	9	3	3	-	-	-	-	-	-	-	-	-	-	-	-	0	0	2	2	16	27	18	24
SS V	3	-	-	9	2	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	2	1	16	18	12	24
SS VIII	-	2	-	16	3	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	2	6	150	48	60	120
SS IX	-	2	-	16	5	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	2	2	30	80	40	60
SS X	-	-	2	6	2	-	-	-	-	-	-	-	-	-	-	-	-	-	2	2	3	3	75	12	30	30
SS XI	-	-	1	7	7	1	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	10	10	49	7	70
SS XII	-	-	1	7	7	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	3	3	9	49	21	21
SS XIII	-	-	1	7	7	1	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	11	11	49	7	77
SS XIV	-	-	1	7	7	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	3	1	12	49	21	28
SS XV	-	-	1	7	7	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	1	16	49	28	28
HS II	3	-	-	9	3	-	-	1	4	-	-	-	-	-	-	-	-	-	1	1	1	3	18	27	36	36
HS III	3	-	-	9	3	-	-	1	5	-	-	-	-	-	-	-	-	-	1	1	1	3	60	27	36	45
HS IV	-	-	2	6	3	-	-	2	12	-	-	-	-	-	-	-	-	-	1	1	1	3	45	18	27	30
RR II	-	-	-	2	4	4	-	-	-	11	7	3	1	2	-	-	-	-	2	1	1	1	33	8	12	22
RR III	-	-	-	2	9	-	-	-	-		-	-	1	3	1	1	-	-	18	15	2	2	6272	18	336	336
RR IV	-	-	-	4	2	2	-	-	-	6	3	1	-	-	-	-	-	-	1	1	1	1	15	8	10	36
RR V	-	-	-	4	2	-	-	-	-	-	-	-	-	-	-	-	-	-	2	2	2	2	30	8	20	20
RR VI	-	-	-	2	4	2	1	-	-	9	3	3	-	-	-	-	3	-	2	1	1	1	27	8	12	18
RR VII	-	-	-	2	6	4	-	-	-	-	-	-	1	7	-	-	5	-	3	3	3	1	150	12	40	45
RR VIII	-	-	-	2	5	4	3	-	-	-	-	-	-	-	-	-	5	5	3	1	1	2	50	12	15	40
RH II	-	-	-	2	6	-	-	3	5	-	-	-	1	2	-	-	-	-	1	2	1	1	50	12	20	30
RH III	-	-	-	4	2	-	-	1	3	-	-	-	-	-	-	-	-	-	1	2	2	1	30	8	12	20
RH IV	-	-	-	3	6	2	-	3	5	-	-	-	-	-	-	-	5	-	1	2	1	1	50	18	20	45
RS II	3	-	-	9	3	9	-	-	-	22	1	7	-	-	-	-	-	-	1	7	3	3	261	27	36	193
RS III	3	-	-	9	1	9	-	-	-	17	8	5	-	-	-	-	-	-	1	5	3	3	201	27	36	153
RS IV	-	-	2	6	2	6	-	-	-	9	5	4	-	-	-	-	-	-	2	1	1	3	45	12	18	30
RS V	-	-	1	7	2	7	-	-	-	17	6	5	-	-	-	-	-	-	0	2	1	1	17	14	7	34
RS VI	-	-	1	7	2	7	-	-	-	25	6	12	-	-	-	-	-	-	0	3	1	1	25	14	7	50
RS VIII	3	-	-	9	9	-	-	-	-	-	-	-	8	3	-	-	-	-	1	9	1	13	1456	81	156	756
RS IX	3	-	-	9	9	-	-	-	-	-	-	-	8	3	-	-	-	-	1	18	2	13	2912	81	156	1512
RS X	-	-	2	6	4	-	-	-	-	-	-	-	5	2	-	-	-	-	2	1	1	7	105	21	36	70
RS XI	-	-	1	7	11	-	-	-	-	-	-	-	6	7	-	-	-	-	0	735	3	13	66921	98	91	72072
RS XII	-	-	1	7	11	-	-	-	-	-	-	-	6	7	-	-	-	-	0	980	4	13	89232	98	91	96096
RS XIV	-	2	-	16	8	-	-	-	-	-	-	-	-	-	-	-	-	-	3	17	12	20	6900	128	400	2208
RS XV	-	2	-	16	8	-	-	-	-	-	-	-	-	-	-	-	-	-	3	5	4	20	2300	128	400	736
RS XVI	-	-	5	5	1	-	-	-	-	-	-	-	-	-	-	-	-	-	5	5	1	12	921	32	168	176
RS XVIII	3	-	-	9	5	3	-	-	-	-	-	-	-	-	-	-	2	-	1	1	2	1	64	45	48	60
RS XIX	3	-	-	9	5	3	-	-	-	-	-	-	-	-	-	-	2	-	1	0	1	1	32	45	30	48
RS XX	-	-	2	6	4	2	-	-	-	-	-	-	-	-	-	-	3	-	2	1	1	5	75	24	36	50
RS XXI	-	-	1	7	10	7	-	-	-	-	-	-	-	-	-	-	3	-	1	0	1	11	33	70	30	77
RS XXII	-	-	1	7	9	7	-	-	-	-	-	-	-	-	-	-	2	-	1	0	2	10	40	63	36	70

TABLE VII

Useful designs in family TT

Design	n	k	k'	r	a	R	A	r'	a'	R'	A'	p	p'	q	q'	Rep.	Tr.	Rows	Cols
TT 1	5	4	5	2	1	0	0	3	1	9	1	2	0	2	12	12	10	5	24
TT 2	5	4	5	2	1	0	0	3	4	0	0	2	0	2	12	12	10	5	24
TT 3	5	5	6	3	1	0	0	3	1	0	0	0	0	3	3	3	10	5	6
TT 4	5	5	6	3	1	0	0	3	2	9	1	2	0	3	15	15	10	6	25
TT 5	5	5	6	3	1	0	0	3	5	0	0	2	0	3	15	15	10	6	25
TT 6	5	5	6	3	1	0	0	6	1	9	1	2	0	3	15	15	10	6	25
TT 7	6	5	6	2	1	0	0	4	2	0	0	1	0	2	8	8	15	6	20
TT 8	6	6	10	4	1	0	0	4	1	0	0	0	0	4	4	4	15	6	10

The useful designs in family TT are given in Table VII, where it is necessary to include the auxiliary parameters r, a, R, A, r', a', R', A', to differentiate between designs otherwise identical. It will be seen that in all of them the basic triangular designs are singly linked blocks (SLB), $r = 2$, $k = n - 1$, or their complement, $r = n - 2$, $k = (n - 1)(n - 2)/2$, one way and doubly linked blocks (DLB), $r = n - 2$, $k = n$, the other way.

These values of r and k give rise to the following triangular designs:

SLB $$\frac{q(pn + 2)}{2}, \frac{n(n - 1)}{2}, \frac{qn}{2}, \frac{(pn + 2)(n - 1)}{2}$$

Complement of SLB $$\frac{q(pn + n - 2)}{n - 2}, \frac{n(n - 1)}{2}, \frac{qn}{n - 2}, \frac{(pn + n - 2)(n - 1)}{2}$$

DLB $$\frac{q(pn - p + 2)}{2}, \frac{n(n - 1)}{2}, \frac{q(n - 1)}{2}, \frac{n(pn - p + 2)}{2}$$

Thus, for an O:PP design to be SLB both ways or SLB one way and its complement the other $qn = (p'n + 2)(n - 1)$ or $qn = (p'n + n - 2)(n - 1)$ respectively; it is easily seen that neither of these equations has any integral solutions for $n > 2$. If an O:PP design is the complement of SLB both ways then $2qn = (p'n + n - 2)(n - 1)(n - 2)$. n has no factor in common with $(n - 1)$ nor, if odd, with $(n - 2)$; thus $n/2$ is integral, $= x$ say, and $qx = (p'x + x - 1)(2x - 1)(x - 1)$, which is impossible for $x > 1$, i.e., for $n > 2$. For an O:PP design to be DLB both ways $q(n - 1) = n(p'n - p' + 2)$, which is impossible for $n > 3$.

Thus, no O:PP designs can be SLB or its complement both ways, neither can they be DLB both ways, but there is no reason why designs of either kind should not fit with other triangular designs to make an O:PP design.

Latin square designs. The basic Latin square design has r replicates of n^2 treatments on rn^2/k blocks of k plots each but, as for the triangular design, it can be extended by adding on complete replicates to each block, repeating the whole design and adding on balanced incomplete blocks. Thus, in fully extended form, the basic Latin square design is:

L $$\frac{q(k + pn^2)}{k}, n^2, \frac{qn^2}{k}, k + pn^2, \text{ where } q = ar + AR, \lambda = \frac{R(k - 1)}{n^2 - 1}$$

Thus the Latin square O:PP design has the following parameters:

LL $$\frac{q(k + pn^2)}{k}, n^2, \frac{qn^2}{k}, \frac{q'n^2}{k'}, \text{ where } q = \frac{k(k' + p'n^2)}{n^2}, q' = \frac{k'(k + pn^2)}{n^2}$$

and there are the further restrictions above on q and similarly on q'.

There are only two useful designs in the family LL. Both have $n = 3$, $k = k' = 6$, $p = p' = 0$ and so $q = q' = 4$ while $a = a' = 1$, $r = r' = 4$, $A = A' = R = R' = 0$, thus leading to designs with 4 replicates of 9 treatments on 6 rows and columns. The only difference between the designs is that in one

first associates concur three times and second associates twice in both rows and columns, while in the other first associates concur three times in rows and twice in columns and conversely for second associates. Thus, if the designs are LL 1 and LL 2 respectively, then, using the notation previously adopted (Freeman, 1957b), in LL 1

$$\lambda_1 = \mu_1 = 3, \lambda_2 = \mu_2 = 2 \text{ and in LL2 } \lambda_1 = \mu_2 = 3, \lambda_2 = \mu_1 = 2.$$

Since, for LL 2, $\nu_1 = \nu_2 = 30$, in the same notation, the design has equal efficiency for both types of associates, and it is the only useful O:PP design with this property.

Summary. All known families of O:PP designs with two associate-classes are classified, these including all with at least one member of practicable size, i.e., having more than two replicates or treatments, not more than 30 replicates, treatments, rows or columns, and not more than 150 plots in all. The designs within these limits are tabulated in their families and where a family has no practicable design its smallest member is given.

Acknowledgements. I should like to express my thanks to Dr. N. L. Johnson of University College, London, for his advice throughout the preparation of this paper, particularly with regard to the tabulation of the designs.

REFERENCES

J. W. Archbold, and N. L. Johnson, "A method of constructing partially balanced incomplete block designs," *Ann. Math. Stat.*, 27, (1956), pp. 624–32.

R. C. Bose, "On the construction of balanced incomplete block designs," *Ann. Eugen.*, 9, (1939), pp. 353–99.

R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, "Tables of partially balanced designs with two associate classes," *N. Carolina Agric. Expt. Sta. Tech. Bull.* No. 107 (1954).

R. C. Bose, and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, 23, (1952), pp. 367–83.

R. C. Bose, S. S. Shrikhande, and K. N. Bhattacharya, "On the construction of group divisible incomplete block designs," *Ann. Math. Stat.*, 24, (1953), pp. 167–95.

G. H. Freeman, "Some further methods of constructing regular group divisible incomplete block designs," *Ann. Math. Stat.* 28, (1957a), pp. 479–87.

G. H. Freeman, "Some experimental designs of use in changing from one set of treatments to another," Part I. *J. R. Statist. Soc.*, B 19, (1957b), pp. 154–62.

G. H. Freeman, "Some experimental designs of use in changing from one set of treatments to another," Part II. Existence of the designs. *J. R. Statist. Soc.*, B 19, (1957c), pp. 163–5.

T. N. Hoblyn, S. C. Pearce, and G. H. Freeman, "Some considerations in the design of successive experiments in fruit plantations," *Biometrics*, 10, (1954), pp. 503–15.

MOST ECONOMICAL MULTIPLE-DECISION RULES[1]

By Wm. Jackson Hall

University of North Carolina

0. Summary. This paper is concerned with non-sequential multiple-decision procedures for which the sample size is a minimum subject to either (1) lower bounds on the probabilities of making correct decisions or (2) upper bounds on the probabilities of making incorrect decisions. Such decision procedures are obtained by constructing artificial decision problems for which the minimax strategies provide solutions to problems (1) and (2). These are shown to be "likelihood ratio" and "unlikelihood ratio" decision rules, respectively. Thus, although problems (1) and (2) are formulated in the spirit of the classical Neyman-Pearson approach to two-decision problems, minimax theory is used as a tool for their solution.

Problems of both "simple" and "composite" discrimination are considered and some examples indicated. (Some multivariate examples are given in [4].) Various properties of the decision rules are derived, and relationships with works of Wald, Lindley, Rao and others are cited

1. Simple discrimination.

A *Formulation of the problem* We are concerned with a sequence $X_1, X_2, \cdots$, of real- or vector-valued, independent, and identically distributed random variables, each having a density function f, belonging to some specified class Ω, w.r.t a fixed measure μ.

The decision problem is to formulate a rule for choosing a non-negative integer n (completely non-random), and, after taking an observation

$$x = (x_1, \cdots, x_n)$$

on $X = (X_1, \cdots, X_n)$, for choosing one of m possible alternative decisions $A_1, \cdots, A_m$. A multiple decision rule (m-d r) for choosing among $A_1, \cdots, A_m$ on the basis of x is defined by an ordered set of non-negative, real-valued, measurable functions $\phi(x) = [\phi_1(x), \cdots, \phi_m(x)]$ on the space $\mathfrak{X}$ of x such that $\sum_i \phi_i = 1$ identically in x (for $n = 0$, the ϕ_i's are constants). A_i is then chosen with probability $\phi_i(x)$ when x is observed. For non-randomized d.r.'s (all ϕ_i's equal 0 or 1), the ϕ_i's are characteristic functions of mutually exclusive and exhaustive "acceptance" regions $R_1, \cdots, R_m$ in $\mathfrak{X}$, where A_i is accepted if $x \in R_i$.

Received January 21, 1956; revised November 15, 1957.

[1] This research was sponsored partly by the Office of Naval Research under Contract No Nonr-855(06) for research in probability and statistics at Chapel Hill and partly by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. 18(600)-458. Reproduction in whole or in part for any purpose of the United States Government is permitted.

A subscript or superscript n denotes the corresponding sample size; $f^n(x)$ and μ^n are the joint density and product measure, respectively.

We suppose throughout Section 1 that Ω consists of a finite number, say l, of elements $f_1, \cdots, f_l$; we say that the corresponding decision problem is one of "simple discrimination" and a d.r. is a d.r. for "discriminating among $f_1, \cdots, f_l$." Here, if μ is non-atomic, only non-randomized d.r.'s need be considered [2].

A d.r. $D = D_n$ is characterized by the functions

$$p_{ij}(D) = \Pr(D \text{ chooses } A_j \mid f_i) = \int_{\mathfrak{X}} \phi_j(x) f_i^n(x)\, d\mu^n \quad (i = 1, \cdots, l; j = 1, \cdots, m).$$

We consider two different criteria for choosing a d.r. for simple discrimination. The first assumes that $l = m$ and that the decision A_i is to be preferred when f_i is true. Denote $p_{ii}(D) = p_i(D) = 1 - q_i(D)$, so that p_i is the probability of a "correct" decision and q_i the probability of an "incorrect" decision when f_i is true.

DEFINITION 1. Let $\alpha = (\alpha_1, \cdots, \alpha_m)$ be a given vector of positive constants each less than one. A d.r. D_N, based on a sample of size N, is said to be a most economical m-decision rule relative to the vector α for discriminating among $f_1, \cdots, f_m$ if it satisfies

$$p_i(D) \geqq \alpha_i \qquad (i = 1, \cdots, m) \tag{1}$$

and if N is the least integer n for which (1) may be satisfied by some m-d.r. D_n based on a sample of size n. N is said to be the most economical sample size.

We now no longer require that $l = m$, but suppose that corresponding to each f_i one or more of the alternatives A_j is preferable, or "correct," when f_i is true.

DEFINITION 2. Let $\beta = (\beta_{ij})$ be a given $l \times m$ matrix of positive constants such that for every i, j pair for which A_j is a correct decision when f_i is true $\beta_{ij} = 1$. A d.r. D_N, based on a sample of size N, is said to be a most economical m-decision rule relative to the matrix β for discriminating among $f_1, \cdots, f_l$ if it satisfies

$$p_{ij}(D) \leqq \beta_{ij} \quad (i = 1, \cdots, l; j = 1, \cdots, m) \tag{2}$$

and if N is the least integer n for which (2) may be satisfied by some m-d.r. D_n based on a sample of size n. N is said to be the most economical sample size.

If $l = m$ and A_i is preferred when f_i is true, then an M.E. d.r. relative to β also controls the probabilities of correct decisions if $\sum_{j \neq i} \beta_{ij} < 1$ for all i.

If $l = m = 2$, both (1) and (2) reduce to upper bounds on the probabilities of the two kinds of error, and Definitions 1 and 2 define an M.E. 2-d.r. as one with minimum sample size subject to these bounds.

It is intuitively clear (and elementary to prove) that a necessary and sufficient condition for the existence of a M.E. m-d.r. relative to any α or β $(l = m)$ is that there exist uniformly consistent sequences of 2-d.r.'s for discriminating between every pair $\omega_i, \omega_j (i \neq j)$ [5].

We shall utilize elements of Wald's theory of decision functions as given in

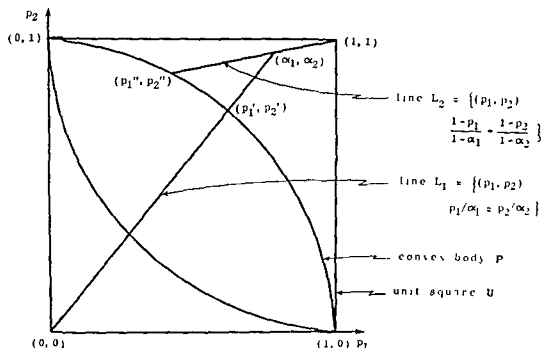


FIG 1

[14], and shall use in particular some of the results of Sections 3.5 and 5.1.1, altering his notation slightly. The differences in the "data of the decision problem" assumed by Wald and here are only minor.

Let $\mathfrak{D}_n$ denote the class of all m-d.r.'s based on a sample of size n ($n = 0, 1, 2, \cdots$). Clearly, for $n \leqq N$, $\mathfrak{D}_n \subseteq \mathfrak{D}_N$; Lemma 1 follows almost immediately.

LEMMA 1. *For every fixed sample size* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. and denote* $r_n = \max_i r(f_i, D_n^0)$, *where* $r(f_i, D_n)$ *is the risk w.r.t. some bounded loss function. Then the sequence* $\{r_n\}$, $n = 0, 1, 2, \cdots$, *is a non-increasing sequence.*

B. *Most economical decision rules relative to a vector* α. Later in this section, we shall apply Wald's theory to two specific loss functions and develop in each case a method of obtaining M.E. d.r.'s as defined by Definition 1. First, we motivate geometrically the selection of such loss functions so as to identify the minimax strategy with the desired one. This alternative approach may give some geometrical insight into the properties of the d.r.'s obtained.[2]

For fixed n, let $p(D) = (p_1(D), \cdots, p_m(D))$ denote a point in m-space, and $P_n = \{p(D): D \varepsilon \mathfrak{D}_n\}$. It can be shown ([2], [10]) that P_n is a convex body in the unit m-cube U containing all corners of U with coordinates summing to unity. The case $m = 2$ is illustrated. (Conditions under which P_n is a proper subset of $P_{n'}$ for $n < n'$ and for which P_n tends with increasing n to U are given elsewhere ([3], [5]).)

---

[2] *The author is indebted to the referee for considerable improvement of this geometric presentation.*

In the diagram, the point $\alpha \not\in P_n$; therefore n is smaller than the required sample size. The M.E. sample size is the smallest n for which $\alpha \in P_n$, in which case the points p', p'' and α coincide (approximately). To test whether or not $\alpha \in P_n$, we can examine the position of the points p' or p'' relative to the position of α.

The points on the "upper" surface of P_n (n fixed) include all points $p(D_n^*)$ corresponding to Bayes strategies D_n^* when the loss function is

$$W(f_i, A_j) = W_{ij} = -\delta_{ij}/\alpha_i \qquad (i, j = 1, \cdots, m), \tag{3}$$

where δ_{ij} denotes the Kronecker δ-function.[3] Then the risk w.r.t. W_{ij} is

$$r(f_i, D) = -\sum_{j=1}^{m} \delta_{ij}\, p_{ij}(D)/\alpha_i = -p_i(D)/\alpha_i \qquad (i = 1, \cdots, m). \tag{4}$$

If p' is not on the boundary of U, the least favorable distribution for the weight function W_{ij} will positively weigh each element in Ω. (This will occur if the region in $\mathfrak{X}$ of positive density is constant over Ω.) In this case, the minimax strategy D_n' will be such that $p(D_n')$ is on the line L_1 of constant risk; i.e., $p(D_n') = p'$. (To obtain the minimax geometry with the loss function W_{ij}, transform the diagram by dividing the ith coordinate by $-\alpha_i$; then the convex body P_n goes into the convex body of risk points $(r_1, \cdots, r_m)$, $r_i = r(f_i, D)$.)

Alternatively, the "upper" surface of P_n corresponds to Bayes strategies when the loss function is

$$W^*(f_i, A_j) = W_{ij}^* = (1 - \delta_{ij})/(1 - \alpha_j) \qquad (i, j = 1, 2, \cdots, m) \tag{5}$$

and the risk function is $r^*(f_i, D) = q_i(D)/\beta_i$, where $\beta_i = 1 - \alpha_i$. The least favorable distribution will likewise positively weight each element of Ω whenever p'' is not on the boundary of U. In this case, the minimax strategy D_n'' will be such that $p(D_n'')$ is on the line L_2 of constant risk; i.e., $p(D_n'') = p''$. (To obtain the minimax geometry with W_{ij}^*, transform the diagram by dividing the ith coordinate by $1 - \alpha_i$, again transforming P_n into the convex body of risk points.) This latter approach is similar to that used by Rao [11] for problems of classification in multivariate analysis.

When $l = m > 2$, there is an added complication for the latter loss function since the line (L_2) from $(1, 1, \cdots, 1)$ through α need not necessarily pierce P for $n < N$, the M.E. sample size. (Of course, if $\alpha \in P_n$, then the line certainly pierces P_n.) Thus the components of a least favorable distribution are not necessarily positive unless $n \geqq N$ and p'' is in the interior of U.

Thus, in one instance, minimax rules maximize the common ratio $p_1/\alpha_1 = \cdots = p_m/\alpha_m$ and, in the other, minimize the common ratio $q_1/\beta_1 = \cdots = q_m/\beta_m$. The M.E. sample size is the smallest one for which the common ratio is $\geqq 1$ or $\leqq 1$, respectively. We now formalize these results. (Wald's Theorem 5.3[4] asserts the existence of a minimax d.r. D^0 for any (fixed) sample size.)

[3] This loss function satisfies Wald's requirements although it is not necessarily zero when a correct decision is made nor necessarily positive otherwise, as intuitively suggested, but never required mathematically, by Wald.

[4] All references to Wald refer to [14] unless otherwise specified.

THEOREM 1. *For each* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. w.r.t. the weight function* (3) *for samples of fixed size* n. *Suppose for some* n,

$$\max_i r(f_i, D_n^0) \leqq -1 \tag{6}$$

and let N *be the least such integer. Then* D_N^0 *is an M.E. d.r. relative to the vector* α *for discriminating among* $f_1, \cdots, f_m$. *Conversely, if there exists an M.E. d.r. relative to* α *for discriminating among* $f_1, \cdots, f_m$, *and the M.E. sample size is* N, *then* D_N^0 *is an M.E. d.r.*

PROOF. From (4) and (6), it follows that D_N^0 satisfies (1). Now suppose for some $n < N$, there exists a d.r. D_n satisfying (1). Since D_n^0 is minimax, $\max_i r(f_i, D_n^0) \leqq \max_i r(f_i, D_n) = \max_i [-p_i(D_n)/\alpha_i]$. Since D_n satisfies (1), we have from above that $\max_i r(f_i, D_n^0) \leqq -1$, in contradiction to the fact that N is the least integer n for which this is true. Hence, D_N^0 is an M.E. d.r.

To prove the converse, suppose D_N is an M.E. d.r. Then

$$-1 \geqq \max_i [-p_i(D_N)/\alpha_i] = \max_i r(f_i, D_N) \geqq \max_i r(f_i, D_N^0)$$

since D_N^0 is a minimax d.r. Hence, (6) is satisfied for $n = N$, and since N is the M.E. sample size, D_N^0 is an M.E. d.r.

Lemma 1 assures us that any n for which (6) is violated is too small. Now let us consider the structure of minimax d.r.'s for a fixed sample size n.

DEFINITION 3. A d.r. D defined by $\phi(x)$ is said to be a likelihood ratio d.r. if there exist positive constants $a_1, \cdots, a_m$ such that for any j and any x for which $\phi_j(x) > 0$, $a_j f_j^n(x) \geqq a_i f_i^n(x)$ for all $i \neq j$.
(Note that $a_1, \cdots, a_m$ determine ϕ completely except in sets of x for which $a_i f_i^n(x) = \max_j a_j f_j^n(x)$ for more than one value of i.) Setting $a_j = \xi_j/\alpha_j$, where $\xi = (\xi_1, \cdots, \xi_m)$ is an a priori distribution over $\Omega = (f_1, \cdots, f_m)$, it follows from Wald's Theorem 5.1 (with (5.6) replaced by (5.7)) that a Bayes d.r. relative to any ξ for which all $\xi_i > 0$ is a likelihood ratio d.r., and conversely.

Wald's Theorem 5.3 asserts the existence of a minimax d.r. and a least favorable distribution, and that any minimax d.r. is a Bayes d.r. relative to any least favorable distribution. Moreover, it follows from (4) and Wald's Theorem 5.3 (iii) that if all components of a least favorable distribution are positive, any minimax d.r. D^0 has the property:

$$p_1(D^0)/\alpha_1 = \cdots = p_m(D^0)/\alpha_m. \tag{7}$$

We shall give sufficient conditions for this to be true.

ASSUMPTION 1. If R is a subset of $\mathfrak{X}$ for which $\int_R f_i^n(x)\, d\mu^n = 0$ for some i, then $\int_R f_i^n(x)\, d\mu^n = 0$ for all values of i. (Whenever this assumption is made, we shall tacitly assume that $\mathfrak{X}$ is redefined so that $f_i^n(x) > 0$ for all i and $x \in \mathfrak{X}$.)

We state a theorem analogous to Wald's Theorem 5.4;[5] the proof (not given) is also analogous.

---

[5] It might be noted that Wald's condition (iii) of Theorem 5.4 is superfluous since it is always fulfilled; e.g., in Wald's notation, let $\delta = 1/u$ $(i = 1, \cdots, u)$ identically in x, and then $r(F_j, \delta) = (u - 1)/u < 1$ for $j = 1, \cdots, k$.

THEOREM 2. *If Assumption 1 holds, all components of a least favorable distribution ξ^0 w.r.t. the weight function w_{ij} are positive.*

Hence, under Assumption 1, an M.E. d.r. may be obtained by the following method: for each sample size n, find a likelihood ratio d.r. D_n^0 for the constants $a_1, \cdots, a_m$ determined by Eqs. (7), and then choose N as the minimum n for which $p_1(D_n^0) \geqq \alpha_1$.

As an alternative approach, we can consider the weight function W_{ij}^* and proceed analogously to the first approach, giving a theorem identical to Theorem 1 with (6) replaced by $\max_i r(f_i, D_n^0) \leqq 1$; and, replacing $a_j = \xi_j/\alpha_j$ by ξ_j/β_j, it follows analogously that a Bayes d.r. relative to any ξ for which all $\xi_i > 0$ is a likelihood ratio d.r., and conversely. Moreover, if all components of a least favorable distribution are positive, any minimax d.r. D^0 has the property:

$$(8) \qquad q_1(D^0)/\beta_1 = \cdots = q_m(D^0)/\beta_m .$$

We shall give sufficient conditions for this to be true. Analogously to Wald's Theorem 5.4, we have:

LEMMA 2. *If Assumption 1 holds, and if there exists some d.r. D for which*

$$r(f_i, D) < 1/\max_{1 \leqq j \leqq m} \beta_j \qquad (i = 1, \cdots, m),$$

then all components of a least favorable distribution are positive.

The following lemma may be useful in this regard:

LEMMA 3. *If $\beta_i < [1/(m-1)] \sum_{j=1}^m \beta_j$ (i.e., $\alpha_i > [\sum \alpha_j - 1]/[m-1]$) for all i, then there exists a d.r. D for which $r(f_i, D) < 1/\max_j \beta_j$ for all i.*

The proof follows by considering a d.r. defined by $\phi_i(x) = 1 - (m-1)\beta_i/\sum \beta_j > 0$ identically in x $(i = 1, \cdots, m)$.

THEOREM 3. *Suppose Assumption 1 holds. For any sample size greater than or equal to the M.E. sample size, all components of a least favorable distribution are positive.*

PROOF. Suppose $n \geqq N$, the M.E. sample size, and that D_n^0 is a minimax d.r. for samples of size n; then, using Lemma 1 and the theorem analogous to Theorem 1, D_n^0 satisfies (1). Use of Lemma 2 completes the proof.

Hence, under Assumption 1, D_N^0 is a likelihood ratio d.r., and an M.E. d.r. may be obtained by considering likelihood ratio d.r.'s D_n^0 for each n for constants $a_1, \cdots, a_m$ determined by (8), and then choosing N as the minimum n for which $q_1(D_n^0) \leqq \beta_1$. If for some n one of the components of a least favorable distribution is zero, we know that n is less than the M.E. sample size (Lemma 1).

A Bayes d.r. relative to any ξ of which all components are positive is admissible [15]. Hence, any likelihood ratio d.r. is admissible, and under Assumption 1 M.E. d.r.'s obtained by either of the above approaches are admissible. Thus, denoting an M.E. d.r. by D_N^0, there does not exist a d.r. D_N' for which $p_i(D_N') \geqq p_i(D_N^0)$ $(i = 1, \cdots, m)$ with strict inequality for at least one i (under Assumption 1).

Suppose now that a real-valued statistic $t = t(x_1, \cdots, x_n)$ exists which is sufficient for the class $\{f_i^n\}$ $(i = 1, \cdots, m)$, and suppose that t has a monotone

likelihood ratio for some ordering of the elements of Ω; i.e., if $g_i(t)$ is the density of t corresponding to $f_i(x)$, then, for some ordering of the subscripts,

$$g_i(t_1)g_j(t_2) \geqq g_i(t_2)g_j(t_1)$$

for $i > j$ and $t_1 > t_2$ [8]. It follows almost immediately that for any $\phi(x)$ which defines a likelihood ratio d.r. there exist constants $\{c_i\}$, $-\infty = c_0 \leqq c_1 \leqq \cdots \leqq c_{m-1} \leqq c_m = \infty$, such that $\phi_i(x) > 0$ implies $c_{i-1} \leqq t(x) \leqq c_i$. Moreover, $\phi_i(x) = 1$ if the latter inequalities are strict, so that randomization may be required only at the points $t = c_i$ and only then if such points have positive probability. Such d.r.'s have been called monotone [1], [8]. If, for example, f_i is of the exponential type $f_i = \beta(\theta_i)e^{\theta_i x}r(x)$, $r \geqq 0$ and θ_i real, for all i, the above conditions are satisfied [1].

Example 1. Suppose f_i is a normal density function with mean θ_i $(-\infty < \theta_1 < \cdots < \theta_m < \infty)$ and known variance σ^2. Then $t = \bar{x}$ is sufficient and the c_i's and N may be obtained by first solving the following equations (iteratively) for the c_i^n's and n with $\rho_n = 1$:

$$(9) \qquad p_i(D_n) = \Phi[\sqrt{n}(c_i^n - \theta_i)/\sigma] - \Phi[\sqrt{n}(c_{i-1}^n - \theta_i)/\sigma] = \alpha_i\rho_n \qquad (i = 1, \cdots, n),$$

where Φ denotes the standard normal distribution function, and then, choosing N to be the least integer $\geqq n$, re-solving for the c_i's and ρ_N. Such a monotone rule will be minimax w.r.t. W_{ij} for the M.E. sample size. Alternatively, (9) may be replaced by equations of the form $1 - p_i(D_n) = (1 - \alpha_i)\rho_n'$, and a solution obtained which will be minimax w.r.t. W_{ij}^*.

Other examples may be treated analogously, allowing for randomization in the discrete cases if desired.

C. *Most economical decision rules relative to a matrix* β. To obtain M.E. d.r.'s as defined by Definition 2, we shall construct an artificial decision problem whose minimax solution will have the properties desired. For convenience, we replace each β_{ij} which is equal to unity by $+\infty$.

Suppose n fixed, and let Ω' be a set of density functions g_{ij} w.r.t. μ $(i = 1, \cdots, l; j = 1, \cdots, m)$, where $g_{ij} = f_i$ identically in x. Define a weight function $W(g_{ij}, A_k) = W_{ijk}$, where

$$(10) \qquad W_{ijk} = 1/\beta_{ij} \quad \text{if } j = k \ (i = 1, \cdots, l; j, k = 1, \cdots, m) \text{ and } 0 \text{ otherwise.}$$

We consider the artificial decision problem of choosing among $A_1, \cdots, A_m$ when one of the $l' = lm$ density functions g_{ij} is "true", and where the "loss" incurred by choosing A_k when g_{ij} is "true", is $W(g_{ij}, A_k)$. The risk function is $r(g_{ij}, D) = \sum_k W_{ijk}p'_{ijk}(D)$, where $p'_{ijk}(D) = \Pr\ (D \text{ chooses } A_k \mid g_{ij}) = p_{ik}(D)$; thus $r(g_{ij}, D) = p_{ij}(D)/\beta_{ij}$ $(i = 1, \cdots, l;\ j = 1, \cdots, m)$. Wald's Theorem 5.3 asserts the existence of minimax d.r.'s.

THEOREM 4. *For each* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. w.r.t. the weight function* (10) *for discriminating among* $g_{11}, g_{12}, \cdots, g_{lm}$ *for samples of fixed size* n. *Suppose for some* n, $\max_{i,j} r(g_{ij}, D_n^0) \leqq 1$, *and let* N *be the least such integer.*

Then D_N^0 is an M.E. d.r. relative to the matrix β for discriminating among $f_1, \cdots, f_l$. Conversely, if there exists an M.E. d.r. relative to β and N is the M.E. sample size, then D_N^0 is an M.E. d.r.

The theorem may be proved in a similar manner to Theorem 1. Now let us consider the structure of these minimax solutions w.r.t. W_{ijk}.

DEFINITION 4. A d.r. D defined by $\phi(x)$ is said to be an unlikelihood ratio d.r. if there exist non-negative constants a_{ij} ($i \neq j$; $i = 1, \cdots, l$; $j = 1, \cdots, m$), where for each i at least one $a_{ij} > 0$, such that for any k and any x for which $\phi_k(x) > 0$, $\sum_{i \neq k} a_{ik} f_i^n(x) \leqq \sum_{i \neq j} a_{ij} f_i^n(x)$ for all $j \neq k$.

Setting $a_{ij} = \xi_{ij} / \beta_{ij}$, where $\xi = (\xi_{11}, \xi_{12}, \cdots, \xi_{lm})$ denotes an a priori distribution over Ω', we have from Wald's Theorem 5.1 that any Bayes d.r. relative to ξ is an unlikelihood ratio d.r. and conversely. Lindley [10] introduced such d.r.'s, obtained by his "method of minimum unlikelihood." Hereafter, we shall suppose $\xi_{ij} = 0$ for every i, j for which $\beta_{ij} = \infty$ without loss of generality.

Wald's Theorem 5.3 asserts the existence of a least favorable distribution ξ^0, and that any Bayes d.r. relative to ξ^0 is a minimax d.r. and conversely; moreover,

$$p_{ij}(D^0)/\beta_{ij} = \max_{i,j} [p_{ij}(D^0)/\beta_{ij}] \qquad \text{for any } i, j \text{ for which } \xi_{ij}^0 > 0. \tag{11}$$

Apparently, however, there are no general conditions under which all $\xi_{ij}^0 > 0$, and consequently we have no proof of the admissibility of a minimax d.r. In fact, supposing $l = m$ and the β_{ij}'s satisfy $\sum_j^m \beta_{ij} = 1$ for every i, then $\xi_{ij}^0 > 0$ for all i, j would imply $p_{ij} = \beta_{ij}$, regardless of the sample size! Geometrically, the convex body in the $l \cdot m$-dimensional space with coordinate axes p_{ij}, corresponding to all possible d.r.'s for a fixed sample size, is not necessarily intersected by the line determined by $p_{ij}/\beta_{ij} = p_{i'j'} / \beta_{i'j'}$ for all pairs of subscripts corresponding to incorrect decisions. However, we do have the following theorem in this regard, assuming $l = m$ and A_i is "correct" when f_i is true ($i = 1, \cdots, m$).

THEOREM 5. *Suppose Assumption 1 holds and that $\sum_{j \neq i} \beta_{ij} < 1$ for every i. For any sample size greater than or equal to the M.E. sample size, a least favorable distribution ξ^0 has the property $\sum_{i=1}^m \xi_{ij}^0 > 0$ for every j.*

The theorem may be proved by a contradiction, using Assumption 1, Definition 4, Lemma 1, and constructing a Bayes d.r. relative to ξ^0. From Theorem 5 and (11), it follows that $p_{ij}(D_N^0)/\beta_{ij}$ attains its maximum for at least one value of i for every j, where D_N^0 is a minimax d.r. for samples of the M.E. size.

Example 2. We shall consider unlikelihood ratio d.r.'s for samples of size n for Example 1 above. For simplicity, suppose $\sigma = 1$, $l = m = 3$, and $\theta_2 = 0$, the alternatives A_1, A_2, A_3 corresponding respectively to the densities f_1, f_2, f_3.

A d.r. with acceptance regions

$$R_1^n = \{x : h_1^n(x) \leqq h_2^n(x), h_1^n(x), \leqq h_3^n(x)\},$$
$$R_2^n = \{x : h_2^n(x) < h_1^n(x), h_2^n(x) \leqq h_3^n(x)\},$$
$$R_3^n = \{x : h_3^n(x) < h_1^n(x), h_3^n(x) < h_2^n(x)\},$$

where $h_i^n(x) = a_{ji}f_j^n + a_{ki}f_k^n$ and (i, j, k) is a permutation of (1, 2, 3), is an unlikelihood ratio d.r. for the weights (a_{ij}). Denoting the sample mean by $\bar{x}$, we may replace h_i^n by $g_i^n = [a_{ji} \exp(n\theta_j\bar{x} - n\theta_j^2/2) + a_{ki} \exp(n\theta_k\bar{x} - n\theta_k^2/2)]$. Now g_1^n is an increasing function of $\bar{x}$ and g_3^n a decreasing function; g_2^n has a single stationary point, a minimum. By sketching the three g_i^n functions, it is clear that if none of the acceptance regions is to be empty, one of three possibilities must obtain: the acceptance regions are of the form $R_1 = \{x:\bar{x} \leqq c_1 \text{ or } c_3 \leqq \bar{x} \leqq c_4\}$, $R_2 = \{x:c_2 \leqq \bar{x} \leqq c_3\}$, $R_3 = \{x:c_1 \leqq \bar{x} \leqq c_2 \text{ or } \bar{x} \geqq c_4\}$, where either $c_1 = c_2$, or $c_3 = c_4$, or both. (Equality signs have been assigned everywhere in the R_i's for simplicity.) Let c (=2 or 3) denote the number of c_i's to be determined. The c_i's may be obtained by solving $c + 1$ of the six equations $p_{ij} = \rho\beta_{ij}$ for the c_i's and ρ, the choice of the equations to be solved being such that $p_{ij} \leqq \rho\beta_{ij}$ for all six pairs of subscripts. Theorem 5 may be helpful in this choice of equations. To obtain an M.E. d r, the sample size n is to be minimized subject to $\rho = \rho_n \leqq 1$. Similar methods may be applied to simple discrimination problems concerning any distribution of the exponential type

2. Composite discrimination.

A. *The problem.* In this section we allow a continuum of possible density functions. For specificity, assume Ω to be the space of a real- or vector-valued parameter θ indexing the class of density functions w.r t μ with elements $f(x, \theta)$.

We further suppose that disjoint subsets $\omega_1, \cdots, \omega_l$ of Ω are specified such that for every pair i, j $(i = 1, \cdots, m; j = 1, \cdots, l)$ there is a definite preference for or against the decision A_j if the true $\theta \varepsilon \omega_i$. We suppose that none of the decisions is definitely preferred if θ is not in some ω_i; this "indifference region" is excluded from Ω for convenience. Under these assumptions, we say that the corresponding decision problem is one of "composite discrimination" and a d.r. is a d.r. for "discriminating among $\omega_1, \cdots, \omega_l$." A d r $D = D_n$ is characterized by the functions

$$p_j(\theta, D) = \Pr (D \text{ chooses } A_j|\theta) = \int_{\mathfrak{X}} \phi_j(x)f^n(x, \theta)d\mu^n \quad (j = 1, \cdots, m),$$

defined for all $\theta \varepsilon \Omega$.

We again consider two criteria for choosing a d.r. for composite discrimination The first requires $l = m$ and A_i to be a "correct" decision if $\theta \varepsilon \omega_i$ and "incorrect" if $\theta \varepsilon \omega_j (j \neq i)$. For the second criterion, we suppose that corresponding to each ω_i one or more alternatives A_j is preferable when $\theta \varepsilon \omega_i$.

The definitions and comments of Section 1.A may be restated, substituting only ω_i for f_i, $\inf_{\theta\varepsilon\omega_i} p_i(\theta, D)$ for $p_i(D)$, and $\sup_{\theta\varepsilon\omega_i} q_i(\theta, D)$ for $q_i(D)$. When $l = m = 2$, an M.E. 2-d.r. may be considered as a test of the hypothesis that $\theta \varepsilon \omega_1$ against the class of alternatives $\theta \varepsilon \omega_2$, satisfying bounds on the two kinds of error; such a d.r. may be obtained by considering, for each n, tests of size $1 - \alpha_1$ w.r.t. ω_1, which maximize the minimum power w.r.t. ω_2 and choosing that test for which n is a minimum subject to the minimum power being at least α_2 [7].

Before extending this result to m-d.r.'s for composite discrimination, we require some results in minimax decision theory.

B. *Minimax decision rules for fixed sample sizes.* We prove three theorems which may be useful in finding minimax d.r.'s. Also, if a sufficient statistic with a monotone likelihood ratio exists, Karlin and Rubin's complete class theorem may be applicable [1], [8]. Sverdrup's results [13] should also be noted.

We shall use a number of Wald's results in [14], Section 3.5 and 5.1.4, with some alteration in his assumptions and notation. We denote a weight function by $W(\theta, A_j) = W_j(\theta)$ $(j = 1, \cdots, m)$ and the corresponding risk function when using a d.r. D by $r(\theta, D)$. An a priori distribution over the Borel subsets $\{\omega\}$ of Ω is denoted by $\Xi = (\xi, \lambda)$, where $\Xi(\omega) = \Pr(\theta \varepsilon \omega) = \sum_{i=1}^{l} \xi_i \lambda_i(\omega)$ and $\xi_i = \Xi(\omega_i)$, $\lambda_i(\omega) = \Pr(\theta \varepsilon \omega | \theta \varepsilon \omega_i)$ $(i = 1, \cdots, l)$. The average risk relative to Ξ is denoted by $r(\Xi, D)$. Other terminology and notation will be self-evident. Wald's Assumptions 5.1 and 5.6, his remarks on page 148 characterizing a Bayes solution, and his theorems 5.11, 5.12, 3.8, 3.9, and 3.10 characterizing minimax solutions are especially pertinent to what follows. Lehmann's existence theorem for least favorable distributions [9] might also be noted.

Assumption 2. *For each i, j pair $(i = 1, \cdots, l; j = 1, \cdots, m)$, $W_j(\theta)$ equals a constant, say W_{ij}, for all $\theta \varepsilon \omega_i$.*
(That is, for each alternative, the loss varies only from subset to subset among $\omega_1, \cdots, \omega_l$ and not within any subset.) This assumption is sufficient to imply the validity of Wald's Assumptions 3.1 to 3.6 (see his remarks on page 148). For a given set of conditional distributions $\lambda = (\lambda_1, \cdots, \lambda_l)$, we denote

$$f_i^{\lambda}(x) = \int_{\omega_i} f^n(x, \theta)\, d\lambda_i \qquad (i = 1, \cdots, l); \tag{12}$$

n is fixed and need not be evident in the notation.

Theorem 6. *If Assumption 2 holds, a necessary and sufficient condition for a d.r. D^* to be a Bayes d.r. relative to $\Xi = (\xi, \lambda)$ for discriminating among $\omega_1, \cdots, \omega_l$ is that D^* be a Bayes d.r. relative to ξ for discriminating among $f_1^{\lambda}, \cdots, f_l^{\lambda}$ w.r.t. the weight function W_{ij}. The average risk in the two cases are equal.*

Proof. Using Assumption 2 and (12), we have

$$\int_{\Omega} W_j(\theta) f^n(x, \theta)\, d\Xi = \sum_{i=1}^{l} \xi_i W_{ij} f_i^{\lambda}(x).$$

The first part of the theorem follows immediately, using Wald's Theorem 5.1 and second paragraph on page 148. By expressing $r(\theta, D)$ as in Wald's (5.81), interchanging the order of integration, using (12) and Wald's (5.2), we have for any d.r. D,

$$\int_{\omega_i} r(\theta, D)\, d\lambda_i = r(f_i^{\lambda}, D). \tag{13}$$

Denoting by $r_{\lambda}(\xi, D)$ the average risk relative to ξ when discriminating among $f_1^{\lambda}, \cdots, f_l^{\lambda}$, we have

$$r(\Xi, D) = r_{\lambda}(\xi, D), \tag{14}$$

completing the proof.

THEOREM 7. *Suppose Assumption 2 holds. Necessary and sufficient conditions that* $\Xi^0 = (\xi^0, \lambda^0)$ *be a least favorable distribution and* D^0 *a minimax d.r. for discriminating among* $\omega_1, \cdots, \omega_l$ *are that*

(i) ξ^0 *is a least favorable distribution and* D^0 *is a minimax d.r. w.r.t.* W_i, *for discriminating among* $f_1^{\lambda^0}, \cdots, f_l^{\lambda^0}$; *and*

(ii) *for any* i *for which* $\xi_i^0 > 0$, $\int_{\omega_i} r(\theta, D^0)\, d\lambda_i^0 = \sup_{\theta \in \omega_i} r(\theta, D^0)$. *Moreover, the maximum risks in the two cases are equal; i.e.,*

$$\sup_{\Omega} r(\theta, D^0) = \max_{1 \leq i \leq l} r(f_i^{\lambda^0}, D^0). \tag{15}$$

PROOF. Necessity: Since Ξ^0 is least favorable, $\inf_D r(\Xi^0, D) \geq \inf_D r((\xi, \lambda^0), D)$ for any ξ, so that, using (14), $\inf_D r_{\lambda^0}(\xi^0, D) \geq \inf_D r_{\lambda^0}(\xi, D)$; that is, ξ^0 is least favorable. Using Wald's Theorem 3.9 and then Theorem 6, D^0 is a Bayes d.r relative to Ξ^0 and a minimax d.r. for discriminating among $f_1^{\lambda^0}, \cdots, f_l^{\lambda^0}$.

We shall now verify (15). Using Wald's Theorem 5.3 (iii), $\max_i r(f_i^{\lambda^0}, D^0) = \sum_i \xi_i^0 r(f_i^{\lambda^0}, D^0) = r_{\lambda^0}(\xi^0, D^0)$, so that together with (14) and Wald's Theorem 3 10, we have $\max_i r(f_i^{\lambda^0}, D^0) = r(\Xi^0, D^0) = \sup_\Omega r(\theta, D^0)$. Continuing with the necessity, for any i for which $\xi_i^0 > 0$, we have $r(f_i^{\lambda^0}, D^0) = \max_i r(f_i^{\lambda^0}, D^0)$ and $\sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$ by Wald's Theorem 3.10, which, together with (15) and (13), prove (ii).

Sufficiency: By Wald's Theorem 3.9 and Theorem 6, D^0 is a Bayes d.r. relative to $\Xi^0 = (\xi^0, \lambda^0)$; i.e., $r(\Xi^0, D^0) = \inf_D r(\Xi^0, D)$. Hence, we need only prove that Ξ^0 is a least favorable distribution. Suppose it is not; then there exists a $\Xi = (\xi, \lambda)$ such that $\inf_D r(\Xi^0, D) < \inf_D r(\Xi, D)$. But $\inf_D r(\Xi, D) \leq r(\Xi, D^0) = \sum_i \xi_i \int_{\omega_i} r(\theta, D^0)\, d\lambda_i \leq \sum_i \xi_i \sup_{\omega_i} r(\theta, D^0) \leq \sup_\Omega r(\theta, D^0)$. Combining these last three results, $r(\Xi^0, D^0) < \sup_\Omega r(\theta, D^0)$.

By Wald's Theorem 5.3 (iii), for any i for which $\xi_i^0 > 0$, $r(f_i^{\lambda^0}, D^0) = \max_i r(f_i^{\lambda^0}, D^0)$, which, together with (13) and (ii), implies $\sup_{\omega_i} r(\theta, D^0) = \max_i \sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$. Hence, from (ii),

$$r(\Xi^0, D^0) = \sum_i \xi_i^0 \int_{\omega_i} r(\theta, D^0)\, d\lambda_i^0 = \sup_\Omega r(\theta, D^0),$$

a contradiction. Q.E.D.

THEOREM 8. *Suppose Assumption 2 holds, and suppose* $\{\lambda^\nu\}$ *is a sequence of sets of conditional a priori distributions and* D^0 *a d.r. such that*

$$\lim_{\nu \to \infty} \int_{\omega_i} r(\theta, D^\nu)\, d\lambda_i^\nu = \sup_{\omega_i} r(\theta, D^0) \qquad (i = 1, \cdots, l), \tag{16}$$

where for each $\nu = 1, 2, \cdots$, D^ν *is a minimax d.r. for discriminating among* $f_1^{\lambda^\nu}, \cdots, f_l^{\lambda^\nu}$. *Then* D^0 *is a minimax d.r. for discriminating among* $\omega_1, \cdots, \omega_l$.

PROOF. By Wald's Theorem 5.3, for each ν there exists a least favorable distribution ξ^ν, and D^ν is a Bayes d.r. relative to ξ^ν for discriminating among $f_1^{\lambda^\nu}, \cdots, f_l^{\lambda^\nu}$; i.e., for any d.r. D, $r_{\lambda^\nu}(\xi^\nu, D^\nu) \leq r_{\lambda^\nu}(\xi^\nu, D)$, and hence, using (14),

$$\sum_i \xi_i^\nu \int_{\omega_i} r(\theta, D^\nu)\, d\lambda_i^\nu \leq \sum_i \xi_i^\nu \int_{\omega_i} r(\theta, D)\, d\lambda_i^\nu \leq \sup_\Omega r(\theta, D) \tag{17}$$

Now each sequence $\{\xi_i^\nu\}$ has at least one limit point; let $\{\Xi^{\nu_j}\}$, $j = 1, 2, \cdots$, be a sub-sequence of $\{\Xi^\nu = (\xi^\nu, \lambda^\nu)\}$ for which each $\xi_i^{\nu_j}$ converges to a limit, say ξ_i^0; then $\sum_i \xi_i^0 = 1$. By Wald's Theorem 5.3 (iii) and (13), for each i for which $\xi_i^\nu > 0$, $\int_{\omega_i} r(\theta, D^\nu)\, d\lambda_i^\nu = \max_i \int_{\omega_i} r(\theta, D^\nu)\, d\lambda_i^\nu$ so that, from (16), for each i for which $\xi_i^0 > 0$, $\sup_{\omega_i} r(\theta, D^0) = \max_i \sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$. Hence, from (16), $\lim_{j=\infty} \sum_i \xi_i^{\nu_j} \int_{\omega_i} r(\theta, D^{\nu_j})\, d\lambda_i^{\nu_j} = \sum_i \xi_i^0 \sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$, which, together with (22), asserts $\sup_\Omega r(\theta, D^0) \leqq \sup_\Omega r(\theta, D)$ for any D. Q.E.D.

If a least favorable distribution exists, the problem reduces to one of simple discrimination, so that if μ is non-atomic only non-randomized d.r.'s need be considered. A lemma for the case of composite discrimination analogous to Lemma 1 may be derived.

C. *Most economical decision rules relative to a vector* α. As in Section 1.B, we shall apply the above theory to two specific weight functions $W_j(\theta)$ and develop in each case a method of obtaining M.E. d.r.'s relative to α. We assume $l = m$. First, let

$$(18) \qquad W(\theta, A_j) = W_j(\theta) = -1/\alpha_j \qquad \text{if } \theta \varepsilon \omega_j \text{ and } 0 \text{ otherwise.}^6$$

The risk w.r.t. $W_j(\theta)$ is $r(\theta, D) = -p_i(\theta, D)/\alpha_i$ if $\theta \varepsilon \omega_i$ $(i = 1, \cdots, m)$, and $\sup_{\omega_i} r(\theta, D) = -\inf_{\omega_i} p_i(\theta, D)/\alpha_i$ $(i = 1, \cdots, m)$. By Wald's Theorem 5.12 (i), there exists a minimax d.r. D^0 for any (fixed) sample size.

THEOREM 9. *For each $n = 0, 1, 2, \cdots$, let D_n^0 be a minimax d.r. w.r.t. the weight function* (18) *for samples of fixed size n. Suppose for some n, $\sup_\Omega r(\theta, D_n^0) \leqq -1$ and let N be the least such integer. Then D_n^0 is an M.E. d.r. relative to α for discriminating among $\omega_1, \cdots, \omega_m$. Conversely, if there exists an M.E. d.r. relative to α for discriminating among $\omega_1, \cdots, \omega_m$, and the M.E. sample size is N, then D_N^0 is an M.E. d.r.*

The proof is like that of Theorem 1, replacing $p_i(D_n)$ by $\inf_{\omega_i} p_i(\theta, D_n)$.

Note that (18) satisfies Assumption 2 with W_{ij} given by (3). Hence, if a least favorable distribution $\Xi^0 = (\xi^0, \lambda^0)$ exists, Theorems 6 and 7 imply that the composite discrimination problem may be treated as a simple discrimination problem with $f_i(x) = f_i^{\lambda^0}(x) = \int_{\omega_i} f(x, \theta)\, d\lambda_i^0$, and the theory of Section 1 will be applicable. If a least favorable distribution does not exist, Theorem 8 asserts that by a similar treatment for a sequence of a priori distributions having certain properties in the limit, it may be possible to solve the composite discrimination problem. Now suppose a least favorable distribution $\Xi^0 = (\xi^0, \lambda^0)$ exists. By Theorem 7(ii),

$$\int_{\omega_i} p_i(\theta, D^0)\, d\lambda_i^0 = \inf_{\theta \varepsilon \omega_i} p_i(\theta, D^0) \qquad \text{for any } i \text{ for which } \xi_i^0 > 0.$$

ASSUMPTION 3. If R is a subset of $\mathfrak{X}$ for which $\int_R f^n(x, \theta)\, d\mu^n = 0$ for some $\theta \varepsilon \Omega$, then $\int_R f^n(x, \theta)\, d\mu^n = 0$ for all $\theta \varepsilon \Omega$.

This assumption implies Assumption 1 for the density functions $f_1^\lambda, \cdots, f_m^\lambda$,

[6] See footnote 3.

defined by (12), for any set of conditional distributions λ. If Assumption 3 holds, and if a least favorable distribution exists, it follows from Theorem 2, Wald's Theorem 5 3(iii) and (18) that

$$(19) \qquad \frac{1}{\alpha_1} \inf_{\theta \varepsilon \omega_1} p_1(\theta, D^0) = \cdots = \frac{1}{\alpha_m} \inf_{\theta \varepsilon \omega_m} p_m(\theta, D^0),$$

where D^0 is a minimax d r.

As a second approach, consider the weight function:

$$(20) \qquad W(\theta, A_j) = W_j(\theta) = 1/\beta_i \qquad \text{if } \theta \varepsilon \omega_i, i \neq j, \text{ and 0 otherwise,}$$

where $\beta_i = 1 - \alpha_i$ as before. Then $r(\theta, D) = q_i(\theta, D)/\beta_i$ if $\theta \varepsilon \omega_i$ $(i = 1, \cdots, m)$. We may proceed analogously to the first approach, making changes corresponding to those made analogously in Section 1. We thus obtain a theorem analogous to Theorem 9 and also

THEOREM 10. *Suppose Assumption 3 holds and that a least favorable distribution exists For any sample size greater than or equal to the M E. sample size,*

$$(21) \qquad \frac{1}{\beta_1} \sup_{\theta \varepsilon \omega_1} q_1(\theta, D^0) = \cdots = \frac{1}{\beta_m} \sup_{\theta \varepsilon \omega_m} q_m(\theta, D^0)$$

where D^0 is a minimax d.r.

No proof of admissibility of the M.E. d r.'s derived in this section has been obtained. However, if Assumption 3 holds and there exists a least favorable distribution, it can easily be verified that there does not exist a d r. D'_N for which $\inf_{\omega_i} p_i(\theta, D'_N) \geqq \inf_{\omega_i} p_i(\theta, D^0_N)$ $(i = 1, \cdots, m)$ with strict inequality for at least one i, where D^0_N is an M.E. d.r. obtained by either of the minimax methods.

D. *Most economical decision rules relative to a matrix β.* Just as the approach of Section 1.B was extended in Section 1.C, we shall extend the approach of Section 2.C in this section to the consideration of M.E d.r.'s for composite discrimination relative to $\beta = (\beta_{ij})$.

Suppose n is fixed, and consider parameter spaces $\Omega_1, \cdots, \Omega_m$, each Ω_j being identical to Ω, and denote $\Omega' = U_j\Omega_j$. For each j, denote the corresponding subsets by $\omega_{1j}, \cdots, \omega_{lj}$. Define a weight function $W(\theta, A_k) = W_k(\theta)$ for $k = 1, \cdots, m$, by

$$(22) \qquad \begin{aligned} &W_k(\theta) = 1/\beta_{ij}, \\ &\qquad \text{if } \theta \varepsilon \omega_{ij} \text{ and } j = k \; (i = 1, \cdots, l; j = 1, \cdots, m), \text{ and 0 otherwise.} \end{aligned}$$

Then $r(\theta, D) = p_j(\theta, D)/\beta_{ij}$ if $\theta \varepsilon \omega_{ij}$. Let Ξ be an a priori distribution over Ω' with components $\xi_{ij} = \Xi(\omega_{ij})$ and $\lambda_{ij}(\omega) = \Pr(\theta \varepsilon \omega \mid \theta \varepsilon \omega_{ij})$. For a given set of λ's, denote

$$(23) \qquad g^{\lambda}_{ij}(x) = \int_{\omega_{ij}} f^n(x, \theta) \, d\lambda_{ij}.$$

Theorem 9 may be restated and proved, substituting only (22) for (18), +1 for −1, and β for α. The theorems of Section 2.B may be applied to obtain mini-

max d.r.'s for composite discrimination w.r.t. the weight function (22) by replacing l in the theorems by $l' = l \cdot m$ and replacing single subscripts i by ij and f_i^λ by g_{ij}^λ. If a least favorable distribution exists, then the composite discrimination problem reduces to a problem of simple discrimination among the "average" density functions g_{ij}^λ defined by (23) w.r.t. a set of "least favorable conditional distributions" λ, and Theorem 5 and the remarks of Section 1.C are applicable. Thus, this method of solution gives unlikelihood ratio d.r.'s as M.E. d.r.'s. If a least favorable distribution does not exist, then a minimax d.r. will be a Bayes d.r. in the wide sense and Theorem 8 may be applicable.

Example 3. Suppose $f(x, \theta)$ is a normal density function with variance σ^2 (known) and mean θ, and

$$\omega_1 = \{\theta : \theta \leqq \theta_1\}, \qquad \omega_2 = \{\theta : \theta_2' \leqq \theta \leqq \theta_2''\}, \qquad \omega_3 = \{\theta : \theta \geqq \theta_3\},$$

for some specified $\theta_1 < \theta_2' \leqq \theta_2'' < \theta_3$. It may be shown that the least favorable conditional distributions over ω_1, ω_2, ω_3 (Theorem 7) assign probability one to θ_1, θ_2, θ_3, where $\theta_2 = \theta_2'$ or θ_2'' determined below. Thus, this example reduces to Example 1. (Karlin and Rubin's results [8] also imply that a minimax rule will be monotone in $\bar{x}$; determining the explicit form of the monotone rule is equivalent to showing that the above distribution is least favorable.)

θ_2 is determined as follows:

$$\text{(24)} \qquad \begin{aligned} \theta_2 &= \theta_2' \quad \text{if} \quad p_2(\theta_2', D') \leqq p_2(\theta_2'', D'), \\ \theta_2 &= \theta_2'' \quad \text{if} \quad p_2(\theta_2'', D'') < p_2(\theta_2', D''), \end{aligned}$$

where D' and D'' are the solutions to the corresponding simple discrimination problems with $\theta_2 = \theta_2'$ or θ_2'' for fixed n. We shall show that such a determination of θ_2 is complete and consistent by showing that if $p_2(\theta_2'', D'') > p_2(\theta_2', D'')$ then $p_2(\theta_2'', D') > p_2(\theta_2', D')$ and conversely. From (9), with either a prime or double-prime on ρ, D, c_1, and c_2, we have $c_1 = \theta_1 + \sigma\Phi^{-1}(\alpha_1\rho)/\sqrt{n}$ and

$$c_2 = \theta_3 + \sigma\Phi^{-1}(1 - \alpha_3\rho)/\sqrt{n},$$

where $\Phi^{-1}(x) = t$ is defined by $\Phi(t) = x$. Substituting in $p_2(\theta, D)$ it becomes clear that it is a decreasing function of ρ for fixed θ. Now $\alpha_2\rho' = p_2(\theta_2', D')$ and $\alpha_2\rho'' = p_2(\theta_2'', D'')$ so that

$$\text{(25)} \qquad \alpha_2(\rho'' - \rho') = p_2(\theta_2'', D'') - p_2(\theta_2', D').$$

Suppose $p_2(\theta_2'', D'') > p_2(\theta_2', D'')$; substituting in (25), it follows that $\rho'' > \rho'$ since p_2 is a decreasing function of ρ. For the same reasons,

$$0 < \alpha_2(\rho'' - \rho') < p_2(\theta_2'', D') - p_2(\theta_2', D').$$

Conversely, in the same manner, if $p_2(\theta_2'', D') > p_2(\theta_2', D')$, then

$$\alpha_2(\rho'' - \rho') > p_2(\theta_2'', D'') - p_2(\theta_2'', D'),$$

and ρ'' must be greater than ρ'; hence,

$$0 < \alpha_2(\rho'' - \rho') < p_2(\theta_2'', D'') - p_2(\theta_2', D'').$$

Other examples with exponential density functions may be treated analogously, and also similar examples for Section 2.D.

Example 4. Now suppose σ is also unknown; denote the mean by μ and replace θ in the ω_i's defined in Example 3 by μ/σ.

Denoting Student's ratio by t and the sample sum of squares by s^2, (t, s) is sufficient for $\theta = (\mu, \sigma)$. If we invoke invariance (under changes in scale), it follows from Blackwell and Girshick's work [1] that a minimax invariant rule must be monotone in t. Theorem 8 8 1 in [1] proves, for the m-decision case as well as the 2-decision case, that invariance is no restriction when discriminating among $\theta_1, \cdots, \theta_m$, where $\theta = \mu/\sigma$. Thus a minimax d.r. for discriminating among θ_1, θ_2, θ_3 is monotone in t. By showing that the risk for a monotone rule is a maximum in ω_i at $\mu/\sigma = \theta_i$ (with θ_2 determined as in Example 3), it will follow that a monotone rule in t, with c_1, c_2 and ρ determined by equations of the form (9) with the Φ's replaced by non-central t distribution functions, will be minimax for discriminating among ω_1, ω_2, ω_3.

Alternatively, this same result may be obtained by an application of our Theorem 8, letting λ_i^ν assign probability one to sets of (μ, σ) in which

$$\mu/\sigma = \theta_i$$

and letting σ^{-2} be distributed as χ^2 with degrees of freedom tending to 0 as $\nu \to \infty$. The details appear in [3], adapted from a 2-d.r. argument by Hoeffding.

Example 5. We shall derive a three-decision extension of the sign test for the median of an arbitrary distribution function by adapting an example of Hoeffding [6]. (See also [12].) Analogously, an M.E. d.r. concerning any quartile of an arbitrary distribution may be derived.

Let Ω be the class of all density functions f w r.t. a fixed measure μ on the real line such that $\mu\{x \leqq 0\} > 0$, $\mu\{x > 0\} > 0$ Denote $\theta(f) = \int_{-\infty}^0 f(x)\,d\mu$. Given $\theta_1, \theta_2', \theta_2'', \theta_3$ $(0 < \theta_1 < \theta_2' \leqq \frac{1}{2} \leqq \theta_2'' < \theta_3 < 1)$, let $\omega_1 = \{f:\theta(f) \leqq \theta_1\}$, $\omega_2 = \{f:\theta_2' \leqq \theta(f) \leqq \theta_2''\}$, $\omega_3 = \{f:\theta(f) \geqq \theta_3\}$. The alternatives A_1, A_2, A_3, corresponding to ω_1, ω_2, ω_3, might be that the median of the unknown distribution is "appreciably" less than zero, "close" to zero, "appreciably" greater than zero, respectively.

Let $f(x, \theta) = \theta^{b(x)}(1 - \theta)^{1-b(x)}/c$ if $x \leqq c$ and 0 otherwise where c is an arbitrary positive constant and $b(x) = 1$ if $x \leqq 0$ and 0 otherwise, and let $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ be a set of conditional distributions over ω_1, ω_2, ω_3, respectively, where λ_i assigns probability 1 to $f(x, \theta_i)$ and where θ_2 is to be determined as in Example 3. It is easily verified that a minimax d.r. D_n (n fixed) for discriminating among f_1^λ, f_2^λ, f_3^λ is monotone in $t(x) = \sum_k b(x_k)$, the number of non-positive observations, with c_1, c_2 and values of ϕ_i when $t = c_1$ or c_2 determined so that $p_i(\theta_i, D_n) = \alpha_i\rho$ $(i = 1, 2, 3)$ for some ρ; and

$$p_1(\theta, D_n) = B(c_1 - 1) + a_1 b(c_1), \qquad p_3(\theta, D_n) = 1 - B(c_2) + (1 - a_2)b(c_2),$$

$$p_2(\theta, D_n) = B(c_2 - 1) + a_2 b(c_2) - B(c_1) + (1 - a_1)b(c_1),$$

where $B = B_{n,\theta}$ and $b = b_{n,\theta}$ denote the binomial distribution function and probability function, respectively, and $a_i = \phi_i(c_i) = 1 - \phi_{i+1}(c_i)$. (It may be shown that D_n defined above is also minimax for discriminating among

$$b_{n,\theta_1}, \quad b_{n,\theta_2}, \quad b_{n,\theta_3}.)$$

This λ may be shown to be least favorable, and an M.E. d.r. may be obtained according to Theorem 9 (see Example 1).

3. Acknowledgments. The author wishes to express his sincere thanks to Professor Wassily Hoeffding, whose stimulating lectures inspired this research, and whose criticisms and suggestions have been invaluable in accomplishing it.

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.
[2] A. Dvoretzky, A. Wald and J. Wolfowitz, "Elimination of randomization in certain statistical decision procedures and zero-sum two-person games," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 1–21.
[3] W. J. Hall, "Most economical multiple-decision rules," Institute of Statistics Mimeograph Series No. 115, University of North Carolina, 1954.
[4] W. J. Hall, "An optimum property of Bechofer's single-sample multiple-decision procedure for ranking means and some extensions," Institute of Statistics Mimeograph Series No. 118, University of North Carolina, 1954.
[5] W. J. Hall, "On uniformly consistent sequences of multiple-decision rules," abstract, *Ann. Math. Stat.*, Vol. 27 (1956), p. 221.
[6] Wassily Hoeffding, " 'Optimum' nonparametric tests," *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 83–92.
[7] Wassily Hoeffding, "Lectures on estimation and testing of hypotheses," University of North Carolina, fall quarter 1952 (unpublished).
[8] S. Karlin and H. Rubin, "The theory of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272–299.
[9] E. L. Lehmann, "On the existence of least favorable distributions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 408–416.
[10] D. V. Lindley, "Statistical inference," *J. Royal Stat. Soc.*, Series B, Vol. 15 (1953), pp. 30–76.
[11] C. R. Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952, p. 311.
[12] Erik Ruist, "Comparison of tests for non-parametric hypotheses," *Arkiv for Matematik*, Vol. 3 (1954), pp. 133–163.
[13] Erling Sverdrup, "Weight functions and minimax procedures in the theory of statistical inference," *Archiv for Mathematik og Naturvidenskab*, Vol. 51 (1952), pp. 1–73.
[14] Abraham Wald, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.
[15] A. Wald and J. Wolfowitz, "Characterization of the minimal complete class of decision functions when the number of distributions and decisions is finite," *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 149–157.

ESTIMATION OF THE MEANS OF DEPENDENT VARIABLES

By Olive Jean Dunn[1]

Statistical Laboratory, Iowa State College

1. Summary. Methods are given for constructing sets of simultaneous confidence intervals for the means of variables which follow a multivariate normal distribution.

In section (3), a set of confidence intervals is obtained for each of two special cases, first when the variances are assumed to be known, and second when the variances are assumed to be equal. These two sets have the property that the confidence is known exactly, rather than merely being bounded below. In the case of known variances, the intervals are of fixed lengths (i.e., the lengths are the same from sample to sample); when the variances are unknown, the intervals are of variable lengths. It may be surprising to note that nothing need be known about the covariances in order to obtain confidence intervals of fixed lengths whose confidence coefficient is exact. These intervals are long, and do not make use of all the information provided by the sample.

Each of sections (4) to (7) considers a different method for obtaining confidence intervals of bounded confidence level. In each section a set of fixed lengths is obtained when the variances are assumed to be known, while a set of variable lengths is obtained when the variances are unknown but equal. In section (5) the set of variable lengths applies to the general multivariate normal distribution, all the other confidence intervals in this paper require some assumption concerning the variances.

In section (8) the sets of intervals are compared on the basis of length. One of the bounded confidence level methods, which has been established only for two or three variables or for an arbitrary number of variables with a special type of correlation matrix, is shown to yield the best possible set. Another of the bounded confidence level methods, whose use is established in general, is shown to be almost as good as the best set for confidence coefficients of practical interest.

It is interesting to notice that intervals with bounded confidence level, are found which are much shorter than the ones whose confidence level is exact. This need not surprise us, however. In the case of just one variable, we might easily find that the 95% confidence intervals for the mean using the t-statistic were shorter on the average than 94% confidence intervals using order statistics. Moreover, since in admitting sets of confidence intervals with bounded confidence level we consider a much broader class of methods, we might almost expect that some of them would give better intervals.

2. Introduction. The problem of estimating the unknown means of dependent variables arises frequently in situations where repeated measurements are made

Received February 15, 1957; revised March 25, 1958.

[1] Now at the University of California at Los Angeles.

on the same individuals, and the assumption of independence is unjustifiable. In biological research, for example, growth data are often obtained with measurements taken on n individuals at k different times; the measurements would be highly correlated. The psychologist might measure n individuals' responses to k different levels of a stimulus; again, a high degree of dependence would be expected. The point estimates chosen for the means would be the same as for independent variables; in this paper we wish to develop simultaneous confidence intervals for the means.

Let $y_1, \cdots, y_k$ be k jointly distributed variables whose means are $\mu_1, \cdots, \mu_k$ respectively. A set of simultaneous confidence intervals for $\mu_1, \cdots, \mu_k$ with confidence coefficient $1 - \alpha$ consists of $2k$ functions of the sample values, say g_i and h_i, $i = 1, 2, \cdots, k$, with the following property: if E_i is the event that the interval g_i to h_i covers μ_i, $i = 1, 2, \cdots, k$, then the probability that E_1, $E_2, \cdots, E_k$ occur simultaneously is greater than or equal to $1 - \alpha$, where $0 < \alpha < 1$. Symbolically,

$$P(E_1E_2, \cdots, E_k) = P(g_1 < \mu_1 < h_1, \cdots, g_k < \mu_k < h_k) \geqq 1 - \alpha.$$

If the inequality sign holds, the set is of bounded confidence level.

Paul G. Hoel has in a recent paper [1] given a method for estimating a mean regression curve and a confidence band for it which is applicable to the situations we have in mind provided one assumes the existence of a polynomial regression curve of a given degree. In this paper we shall assume that the experimenter is actually interested in the regression curve, but is either unwilling to make the necessary laborious calculations or else is unable to make the necessary assumptions concerning its form. He knows that there exist methods for studying linear contrasts among the means, but this is not what he wishes to do. He might indeed decide to make k different 95% confidence intervals, one for each of the k means; this is satisfactory only when he focuses on one individual mean.

We shall assume, then, that he will welcome a set of k confidence intervals, one for each mean, being assured, with a high probability, that such a set covers all k means simultaneously.

Another situation in which such a set of intervals would be useful arises when a regression line, curve, or surface has been fitted, and several predictions are made on the basis of it.

Suppose, for example, that the assumption has been made that the variables x_i are normally distributed with means $\alpha + \beta t_i$ and variances σ^2, and that the maximum likelihood estimate $\hat{\alpha} + \hat{\beta} t_i$ has been calculated from a sample of size m.

At any particular value of t, say t_0, one can obtain a prediction interval for x_0, an observation drawn at random from the x's belonging to t_0, by using the fact that $u_0 = x_0 - \hat{\alpha} - \hat{\beta} t_0$ is normally distributed. But the research worker is cautioned not to do this for more than one value of t, and of course this is exactly what he wishes to do.

If he goes ahead and gets such intervals at k different points, say $t_1^*, \cdots, t_k^*$, he has the same unsatisfactory situation as with repeated tests of significance. The variables $u_i^* = x_i^* - \hat{\alpha} - \hat{\beta} t_i$, where x_i^* is an observation chosen at random

from the x's at $t = t_i^*$, $i = 1, 2, \cdots, k$, are normally distributed and are correlated; thus the methods of this paper may be used to give simultaneous prediction intervals for the points $x_1^*, \cdots, x_k^*$.

3. Confidence regions using independent linear combinations.

3.1. Assuming first known variances, we seek independent linear combinations of the sample values which can be used to give a set of confidence intervals of fixed lengths whose confidence level is exact.

The observations $y_{1j}, y_{2j}, \cdots, y_{kj}, j = 1, \cdots, n$, are a random sample of n observations from $n_k(y_1, \cdots, y_k)$, the multivariate normal distribution with unknown means, $\mu_1, \cdots, \mu_k$, known variances, $\sigma_1^2, \cdots, \sigma_k^2$, and unknown covariances λ_{is}, $i \neq s$.

Let $z_i = \sum_{j=1}^n a_{ji} y_{ij}$, $i = 1, \cdots, k$, with the following restrictions on the a_{ji}

$$(1) \quad \sum_{j=1}^n a_{ji} = 1, \qquad i = 1, \cdots, k$$

$$(2) \quad \sum_{j=1}^n a_{ji} a_{js} = 0, \qquad i \neq s$$

$$(3) \quad \sum_{j=1}^n a_{ji}^2 = c^2, \qquad i = 1, 2, \cdots, k.$$

The means, variances, and covariances of the z_i may then be calculated, remembering that $E(y_{ij} - \mu_i)(y_{sj} - \mu_s) = \lambda_{is}$, but that (since two observations in a random sample are independent) $E(y_{ij} - \mu_i)(y_{sj'} - \mu_s) = 0$ for $j \neq j'$. The means of the z_i are calculated to be μ_i, $i = 1, \cdots, k$, their variances are proportional to $\sigma_1^2, \cdots, \sigma_k^2$, and their covariances are zero.

To determine the a_{ji}, let $A = (a_{ji})$, an $n \times k$ matrix. The columns of A may be considered to be k vectors in an n-dimensional Euclidean space, each with an end fixed at the origin. The three conditions imply (1) that the k vectors have their endpoints on the plane which passes through the unit points on the coordinate axes, $P: \sum_{i=1}^n a_i = 1$; (2) that they be mutually orthogonal, and (3) that their lengths equal c.

If $n \geq k$, the columns of

$$D = \left[\begin{array}{cccc} c & 0 & \cdots & 0 \\ 0 & c & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & c \\ \hline 0 & 0 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & 0 \end{array}\right], \text{ an } n \times k$$

matrix, are k mutually orthogonal vectors of length c whose endp[illegible]
plane

$$P': \frac{a_1}{c} + \cdots + \frac{a_k}{c} + \frac{a_{k+1}}{m_{k+1}} + \cdots + \frac{a_n}{m_n} = 1.$$

The plane P' can be rotated into the plane P provided the distances of the two planes from the origin are equal; this will be true if

$$c^2 = \frac{k}{n - \frac{1}{m_{k+1}^2} - \cdots - \frac{1}{m_n^2}}.$$

To make the lengths of the confidence intervals formed from the z_i as small as possible, c^2 should be minimized. This is accomplished by choosing for P' the plane $\sum_{i=1}^{k} (a_i/c) = 1$; then $c^2 = (k/n)$.

The solution is then $A = BCD$, where B is an $n \times n$ orthogonal matrix whose first column consists of the elements $n^{-\frac{1}{2}}$;

$$C = \left[\begin{array}{ccc|ccc} \frac{1}{\sqrt{k}} & \cdots & \frac{1}{\sqrt{k}} & 0 & \cdots & 0 \\ \cdots & & \cdots & & & \\ \cdots & & \cdots & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & & & & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right], \text{ an } n \times n$$

matrix consisting of zeros except for a $k \times k$ orthogonal matrix in the upper left corner whose first row is $k^{-\frac{1}{2}}, \cdots, k^{-\frac{1}{2}}$; and D is defined as before, with $c = (k/n)$. For C rotates the column vectors of D into vectors whose endpoints lie on the plane $a_1 = n^{-\frac{1}{2}}$. B^{-1} rotates the plane $\sum_{i=1}^{n} a_i = 1$ into the plane $a_1 = n^{-\frac{1}{2}}$, so that B rotates the k mutually orthogonal vectors of length $(k/n)^{\frac{1}{2}}$ into vectors whose endpoints lie on the plane $\sum_{i=1}^{n} a_i = 1$. The problem thus reduces to that of writing down a $k \times k$ orthogonal matrix and an $n \times n$ orthogonal matrix.

The $z_1, \cdots, z_k$ are then independently normally distributed with means μ_i and variances $(k/n)\sigma_i^2$. Thus

$$P\left(z_1 - \sqrt{\frac{k}{n}}\,\sigma_1 c_\alpha < \mu_1 < z_1 + \sqrt{\frac{k}{n}}\,\sigma_1 c_\alpha, \cdots, z_k - \sqrt{\frac{k}{n}}\,\sigma_k c_\alpha < \mu_k < z_k + \sqrt{\frac{k}{n}}\,\sigma_k c_\alpha\right) =$$

$1 - \alpha$, where c_α is defined by

$$N(c_\alpha) = \frac{1 + (1 - \alpha)^{1/k}}{2},$$

with N the cumulative distribution function of the standard normal variable. The set of confidence intervals is $z_i \pm (k/n)^{\frac{1}{2}}\sigma_i c_\alpha$.

3.2. When the variances are unknown but are assumed to be equal, the same method may be used to construct t-variables whose numerators are independent but which have the same denominator, provided $n > k$. Let $\sigma_i^2 = \sigma^2$, $i = 1, \cdots, k$.

Let

$$z_i = \sum_{j=1}^{n} a_{ji} y_{ij}, \qquad i = 1, \cdots, k,$$

and

$$u_m = \sum_{j=1}^{n} b_{jm} y_{rj}, \qquad m = 1, \cdots, n - k,$$

where r is any integer from 1 to k. The choice of r is arbitrary. It may be the same for each u_m, or different r's may be used for the different values of m. The problem is to determine the a_{ji} and b_{jm} so that $z_1, \cdots, z_k, u_1, \cdots, u_{n-k}$ will be independently normally distributed variables with $E(z_i) = \mu_i$, $i = 1, \cdots, k$, $E(u_m) = 0, m = 1, \cdots, n - k, E(z_i - \mu_i)^2 = (k/n)\sigma^2, i = 1, \cdots, k; E(u_m^2) = \sigma^2$, $m = 1, \cdots, n - k$. This will be accomplished provided

(1) $\sum_{j=1}^{n} a_{ji} = 1, i = 1, \cdots, k$, since $E(z_i) = \mu_i \sum_{j=1}^{n} a_{ji} = \mu_i$.

(2) $\sum_{j=1}^{n} a_{ji}^2 = \frac{k}{n}, i = 1, \cdots, k$, since $E(z_i - \mu_i)^2 = \sigma^2 \sum_{j=1}^{n} a_{ji}^2 = \frac{k}{n}\sigma^2$

(3) $\sum_{j=1}^{n} a_{ji}a_{js} = 0, i \neq s$, since $E(z_i - \mu_i)(z_s - \mu_s) = \lambda_{..} \sum_{j=1}^{n} a_{ji}a_{js} = 0$

(4) $\sum_{j=1}^{n} b_{jm} = 0, m = 1, \cdots, n - k$, since $E(u_m) = \mu_m \sum_{j=1}^{n} b_{jm} = 0$

(5) $\sum_{j=1}^{n} b_{jm}^2 = 1, i = 1, \cdots, n - k$, since $E(u_m^2) = \sigma^2 \sum_{j=1}^{n} b_{jm}^2 = \sigma^2$.

(6) $\sum_{j=1}^{n} b_{jm}b_{js} = 0, m \neq s$, since $E(u_m u_s) = E(y_r - \mu_r)(y_r - \mu_r) \sum_{j=1}^{n} b_{jm}b_{js} = 0$

(7) $\sum_{j=1}^{n} a_{ji}b_{jm} = 0, i = 1, \cdots, k; m = 1, \cdots, n - k$, since $E(z_i - \mu_i)(u_m)$

$$= \lambda_{..} \sum_{j=1}^{n} a_{ji}b_{jm} = 0.$$

Thus n mutually orthogonal vectors are needed: k of length (k/n) [illegible] points on the plane $\sum c_i = 1$, and $n - k$ of length one [illegible] plane $\sum c_i = 0$.

Let

[illegible]

n $n \times n$ matrix whose columns are n mutually orthogonal vectors of the needed ngths.

Let

$$C = \left[\begin{array}{ccc|cccc} \frac{1}{\sqrt{k}} & \cdots & \frac{1}{\sqrt{k}} & 0 & & \cdots & 0 \\ \cdots & \cdots & & \cdots & & & \cdots \\ \cdots & \cdots & & 0 & & \cdots & 0 \\ \hline 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \cdots & & & & & & \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{array}\right],$$

$n \times n$ orthogonal matrix which rotates the first k columns of D into vectors ose endpoints lie on the plane $a_1 = n^{-\frac{1}{2}}$ and which leaves the last $n - k$ columns changed.

Let B be an $n \times n$ orthogonal matrix whose first column consists entirely of e elements $n^{-\frac{1}{2}}$. Since B rotates the plane $a_1 = n^{-\frac{1}{2}}$ into the plane $\sum_{i=1}^{n} a_i = 1$, must also rotate the parallel plane $a_1 = 0$ into $\sum_{i=1}^{n} a_i = 0$.

Thus $A = BCD$ is an $n \times n$ matrix whose columns are orthogonal vectors. e first k are of length $(k/n)^{\frac{1}{2}}$ and have endpoints on $\sum_{i=1}^{n} a_i = 1$; the last $n - k$ e of length one and have endpoints on $\sum_{i=1}^{n} a_i = 0$.

Then let

$$t_i = \frac{z_i - \mu_i}{\sqrt{\frac{k}{n(n-k)} \sum_{m=1}^{n-k} u_m^2}}, \qquad i = 1, \cdots, k.$$

ese are k t-variables whose numerators are independent but whose denomina-s are the same. Their frequency function is (see [2]):

$$f_{n-k}(t_1, \cdots, t_k) = \frac{\Gamma\left(\frac{n}{2}\right)}{[\pi(n-k)]^{k/2}\Gamma\left(\frac{n-k}{2}\right)} \left(1 + \frac{\sum_{i=1}^{k} t_i^2}{n-k}\right)^{-(n/2)}$$

c_α is defined by

$$\int_{-c_\alpha}^{c_\alpha} \cdots \int_{-c_\alpha}^{c_\alpha} f_{n-k}(t_1, \cdots, t_k)\, dt_1, \cdots, dt_k = 1 - \alpha,$$

n $P(-c_\alpha < t_1 < c_\alpha, \cdots, -c_\alpha < t_k < c_\alpha) = 1 - \alpha$. Thus an exact set of con-ence intervals of equal but variable lengths is obtained:

$$z_i \pm c_\alpha \sqrt{\frac{k}{n(n-k)} \sum_{m=1}^{n-k} u_m^2}, \qquad i = 1, \cdots, k.$$

4. Intervals of bounded confidence level using the chi-square distribution and telling's T-distribution.

4.1. Known variances. For a sample of size n from the multivariate normal tribution with means $\mu_1, \cdots, \mu_k$ and covariance matrix (λ_{ij}), the expression

$n\sum_{i=1}^{k}\sum_{s=1}^{k}\lambda^{is}(\bar{y}_i - \mu_i)(\bar{y}_s - \mu_s)$ follows a Chi-square distribution with k degrees of freedom. Here λ^{is} denotes an element of the inverse matrix $(\lambda^{is}) = (\lambda_{is})^{-1}$, and $\bar{y}_s$ is the sample mean of the observations on y_s. Then

$$\sum_{i=1}^{k}\sum_{s=1}^{k}\lambda^{is}(\bar{y}_i - \mu_i)(\bar{y}_s - \mu_s) = \frac{c_\alpha}{n},$$

where c_α is defined by $U_k(c_\alpha) = 1 - \alpha$, with U_k the cumulative distribution function of a Chi-square variable with k degrees of freedom. In the parameter space of the $\mu_1, \cdots, \mu_k$, this equation defines an ellipsoid, which will be denoted by E. Then

$$P\left(\sum_{i=1}^{k}\sum_{s=1}^{k}\lambda^{is}(\bar{y}_i - \mu_i)(\bar{y}_s - \mu_s) < \frac{c_\alpha}{n}\right) = P[E \text{ covers } (\mu_1, \cdots, \mu_k)] = 1 - \alpha.$$

To obtain a rectangular confidence region of bounded confidence level, a rectangular parallelepiped, say R, with boundary planes parallel to the coordinate planes in the $\mu_1, \cdots, \mu_k$ space is circumscribed around the ellipsoid E. The boundary planes of R are found to be

$$\mu_i = \bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}}\sqrt{c_\alpha},$$

and are not dependent on the correlations.

Then $P[R \text{ covers } (\mu_1, \cdots, \mu_k)] > P[E \text{ covers } (\mu_1, \cdots, \mu_k)] = 1 - \alpha$, thus giving a set of intervals, $\bar{y}_i \pm (\sigma_i/n^{\frac{1}{2}})c_\alpha^{\frac{1}{2}}$, with $U_k(c_\alpha) = 1 - \alpha$.

4.2. Unknown variances. The same method applies when the variances are unknown and $n > k$, using Hotelling's T-statistic. Here E is the ellipsoid $\sum_{i=1}^{k}\sum_{s=1}^{k} l^{is}(\bar{y}_i - \mu_i)(\bar{y}_s - \mu_s) = c_\alpha^2/n$ where (l^{is}) is the inverse of the matrix (l_{is}) and $l_{is} = \sum_{j=1}^{n}(y_{ij} - \bar{y}_i)(y_{sj} - \bar{y}_s)/(n-1)$, $i = 1, \cdots, k$; $s = 1, \cdots, k$. The boundary planes of R, the circumscribed parallelepiped, are $\mu_i = \bar{y}_i \pm (\hat{\sigma}_i/n^{\frac{1}{2}})c_\alpha$, where $\hat{\sigma}_i = l_{ii}^{\frac{1}{2}}$. For c_α defined by $F(c_\alpha) = 1 - \alpha$, with F the c.d.f. of Hotelling's T, the set of confidence intervals is $\bar{y}_i \pm (\hat{\sigma}_i/n^{\frac{1}{2}})c_\alpha$, $i = 1, 2, \cdots, k$.

It is to be noted that this is the only set of intervals given in this paper for which no assumption has been made concerning the variances. For the other sets, the variances were assumed to be known or else to be unknown but equal.

4.3. More general distribution functions. For n large, T^2 can be assumed to follow a Chi-square distribution with k degrees of freedom, even though the original variables are not normally distributed [3]. A set of confidence intervals for $\mu_1, \cdots, \mu_k$ is then $\bar{y}_i \pm (\hat{\sigma}_i/n^{\frac{1}{2}})c_\alpha^{\frac{1}{2}}$, with c_α the upper α point of the Chi-square distribution with k degrees of freedom.

5. Bounded regions based on linear contrasts. Henry Scheffé [1] obtains simultaneous confidence intervals for the totality of linear contrasts among k means, $\mu_1, \cdots, \mu_k$, using the F distribution. He shows that $P(\hat{\theta} - S\hat{\sigma}_{\hat{\theta}} \leq \theta \leq \hat{\theta} + S\hat{\sigma}_{\hat{\theta}}) = 1 - \alpha$. Here θ is any linear contrast; $S^2 = (k-1)c_\alpha$; c_α is the upper α point of the F distribution with $k - 1$ and ν degrees of freedom; ν is the de-

grees of freedom of the χ^2 variable used in estimating the variance; and P denotes the probability that all such intervals cover their corresponding contrasts.

It can easily be shown that confidence intervals for the totality of linear combinations of $\mu_1, \cdots, \mu_k$ are similarly obtained from $P(\theta - S\hat{\sigma}_{\hat{\theta}} \leqq \hat{\theta} \leqq \theta + S)\hat{\sigma}_{\hat{\theta}} = 1 - \alpha$, where now $S^2 = kc_\alpha$, with c_α the upper α point of the F distribution with k and ν degrees of freedom. Since the k means $\mu_1, \cdots, \mu_k$ are a subset of the linear combinations, confidence intervals for them follow immediately.

5.1. Variances known. If the variables $y_1, \cdots, y_k$ are normally distributed with unknown means $\mu_1, \cdots, \mu_k$, known variances $\sigma_1^2, \cdots, \sigma_k^2$, and unknown correlations, ρ_{ij}, then the χ^2 distribution is used rather than the F distribution, and we have:

$$P\left(\bar{y}_1 - \frac{\sigma_1}{\sqrt{n}}\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \frac{\sigma_1}{\sqrt{n}}\sqrt{c_\alpha}, \cdots, \bar{y}_k - \frac{\sigma_k}{\sqrt{n}}\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \frac{\sigma_k}{\sqrt{n}}\sqrt{c_\alpha}\right) \geqq 1 - \alpha.$$

Here c_α is, as in section 4.1, the upper α point of the χ^2 distribution with k degrees of freedom, and the intervals obtained are the same as those of section 4.1.

5.2. Variances unknown but equal. When the variances are unknown but equal, then as an estimate of σ^2 one may use $\hat{\sigma}_1^2 = \sum_{j=1}^{n}(y_{1j} - \bar{y}_1)^2/(n-1)$. Then

$$P\left(\bar{y}_1 - \sqrt{\frac{k}{n}}\,\hat{\sigma}_1\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\,\hat{\sigma}_1\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\,\hat{\sigma}_1\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \sqrt{\frac{k}{n}}\,\hat{\sigma}_1\sqrt{c_\alpha}\right) \geqq 1 - \alpha,$$

with c_α the upper α point of the F distribution with k and $n - 1$ degrees of freedom. The confidence intervals are $\bar{y}_i \pm (k/n)^{\frac{1}{2}}\hat{\sigma}_1 c_\alpha^{\frac{1}{2}}$.

It may seem unsatisfactory to use only the data from one sample point as an estimate of σ^2; this has been done in order to have a χ^2 variable for the denominator of the F variable.

If one wishes to use a pooled estimate of the variance, $\hat{\sigma}_p^2 = \sum_{i=1}^{k}\hat{\sigma}_i^2/k$, then $\hat{\sigma}_p^2$ no longer has a χ^2 distribution because of the dependence of the variables. It is possible to show, however, that the F distribution may still be used, provided for degrees of freedom one uses k and $n - 1$ (rather than k and $k(n - 1)$). That the degrees of freedom may not be increased may be seen by examining the extreme case when all the correlations are equal to one.

To establish the necessary inequality for using $\hat{\sigma}_p^2$, one may fix $\hat{\sigma}_1, \cdots, \hat{\sigma}_k$ and consider the conditional probability

$$P(\bar{y}_1 - \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} < \mu_k < \bar{y}_1 + \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} \mid \hat{\sigma}_1, \cdots, \hat{\sigma}_k)$$

$$\geqq P\left(\bar{y}_1 - \sqrt{\frac{k}{n}}\frac{\Sigma\hat{\sigma}_i}{k}\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\frac{\Sigma\hat{\sigma}_i}{k}\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\frac{\Sigma\hat{\sigma}_i}{k}\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \sqrt{\frac{k}{n}}\frac{\Sigma\hat{\sigma}_i}{k}\sqrt{c_\alpha} \mid \hat{\sigma}_1, \cdots, \hat{\sigma}_k\right)$$

$$\geqq \sum_{i=1}^{k} P\left(\bar{y}_1 - \sqrt{\frac{k}{n}}\hat{\sigma}_i\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\hat{\sigma}_i\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\hat{\sigma}_i\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \sqrt{\frac{k}{n}}\hat{\sigma}_i\sqrt{c_\alpha} \mid \hat{\sigma}_1, \cdots, \hat{\sigma}_k\right) \Big/ k.$$

Thus for the unconditional probability one has:

$$P\left(\bar{y}_1 - \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha}\right) \geqq 1 - \alpha.$$

6. Regions based on a bonferroni inequality. Confidence regions can be obtained very simply using a Bonferroni inequality [5]. The use of this inequality in a related situation was suggested by E. Paulson [6]

6.1. Variances known. Let $n_k(y_1, \cdots, y_k, \cdot\mu_i, \sigma_i^2, \rho_{ij})$ be the frequency function of k normally distributed variables with means $\mu_1, \cdots, \mu_k$, known variances $\sigma_1^2, \cdots, \sigma_k^2$, and unknown correlations ρ_{ij}. Let $\bar{y}_i$ be the mean of a random sample of size n, $y_{i1}, \cdots, y_{in}$.

Let $z_i = ((\bar{y}_i - \mu_i)n^{\frac{1}{2}})/\sigma_i)$, $i = 1, \cdots, k$. Then the joint frequency function of $z_1, \cdots, z_k$ is $n_k(z_1, \cdots, z_k; 0, 1, \rho_{ij})$, and

$$P(-c < z_1 < c, \cdots, -c < z_k < c) = \int_{-c}^{c} \cdots \int_{-c}^{c} n_k(z_1, \cdots, z_k, 0, 1, \rho_{ij})\, dz_1 \cdots dz_k.$$

Using a Bonferroni inequality, this integral is greater than or equal to $1 - 2k(1 - N(c))$, where N is the c.d.f. of a standard normal variable Setting this expression equal to $1 - \alpha$, c_α may be defined by $N(c_\alpha) = 1 - (\alpha/2k)$ Then

$$P\left(-c_\alpha < \frac{(\bar{y}_1 - \mu_1)\sqrt{n}}{\sigma_1} < c_\alpha, \cdots, -c_\alpha < \frac{(\bar{y}_k - \mu_k)\sqrt{n}}{\sigma_k} < c_\alpha\right) = P[R \text{ covers } (\mu_1, \cdots, \mu_k)] \geqq 1 - \alpha,$$

where R is bounded by

$$\mu_i = \bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} c_\alpha.$$

6.2. Variances unknown but equal. Let y_i, $i = 1, \cdots, k$, have the joint frequency function $n_k(y_1, \cdots, y_k; \mu_i, \sigma^2, \rho_{ij})$, where the variances are unknown but equal. Let $z_i = ((\bar{y}_i - \mu_i)n^{\frac{1}{2}})/\sigma$, $i = 1, \cdots, k$.

We wish to define Student t-variables $t_1, \cdots, t_k$ using $z_1, \cdots, z_k$ in the numerators and using the same Chi-square variable in the denominators. If $u_i = \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2/\sigma^2$, then u_i is a Chi-square variable with $n - 1$ degrees of freedom. Since the u_i are not independent of each other, we choose one, say u_1, to use in all the denominators, rather than use their sum which does not have a Chi-square distribution.

Then

$$t_i = \frac{\sqrt{n-1}\, z_i}{u_1^{1/2}} = \frac{\sqrt{n}(\bar{y}_i - \mu_i)}{\hat{\sigma}_1}, \qquad i = 1, \cdots, k,$$

are Student t-variables with the same denominators. Their distribution function [3] is;

$$f_{n-1}(t_1, \cdots, t_k; \rho_{is}) = \frac{\Gamma\left(\frac{k+n-1}{2}\right)}{(n-1)^{k/2}\pi^{k/2}\Gamma\left(\frac{n-1}{2}\right)} \, |(\rho^{is})|^{1/2}\left[1 + (n-1)^{-1}\sum_{i=1}^{k}\sum_{s=1}^{k}\rho^{is}t_it_s\right]^{-\frac{k+n-1}{2}}$$

where ρ^{is} is an element of $(\rho^{is}) = (\rho_{is})^{-1}$, and $|(\rho^{is})|$ is the determinant of (ρ^{is}).

As in 6.1,

$$P(-c < t_1 < c, \cdots, -c < t_k < c) = \int_{-c}^{c} \cdots \int_{-c}^{c} f_{n-1}(t_1, \cdots, t_k; \rho_{is})\, dt_1 \cdots dt_k \geqq 1 - 2k(1 - H_{n-1}(c)),$$

where H_{n-1} is the c.d.f. of a t-variable with $n - 1$ degrees of freedom.

The set of confidence intervals is then

$$\bar{y}_i \pm \frac{\hat{\sigma}_1}{\sqrt{n}} c_\alpha,$$

where

$$H_{n-1}(c_\alpha) = 1 - \frac{\alpha}{2k},$$

and

$$\hat{\sigma}_1 = \sum_{j=1}^{n} (y_{1j} - \bar{y}_1)^2/(n-1).$$

As in section 5.2, it is possible in these confidence intervals to replace $\hat{\sigma}_1$ by $\hat{\sigma}_p$, the pooled estimate of the variance; $n - 1$ must be retained as the degrees of freedom.

7. Regions with bounded confidence level using inequalities between dependent and independent cases.

7.1. Variances known. For $y_1, \cdots, y_k$ independently normally distributed with unknown means $\mu_1, \cdots, \mu_k$ and known variances, $\sigma_1^2, \cdots, \sigma_k^2$, let x_i be

defined by $x_i = (n^{\frac{1}{2}}(\bar{y}_i - \mu_i))/\sigma_i$, where $\bar{y}_i$ is the mean of the n observations on the ith variable. Then

$$P(-c_\alpha < x_1 < c_\alpha, \cdots, -c_\alpha < x_k < c_\alpha) = \prod_{i=1}^{k} P(-c_\alpha < x_i < c_\alpha) = 1 - \alpha,$$

where c_α is defined by $N(c_\alpha) = \frac{1}{2}[1 + (1 - \alpha)^{1/k}]$, with N the c.d.f. of the univariate normal distribution. The set of simultaneous confidence intervals whose exact confidence level is $1 - \alpha$ is then $\bar{y}_i \pm \sigma_i c_\alpha / n^{\frac{1}{2}}$.

If, now, the $y_1, \cdots, y_k$ are defined as above except that now there may be correlations among them, the same confidence intervals can be used as a set with bounded confidence level, provided it can be proved that

$$P(-c_\alpha < x_1 < c_\alpha, \cdots, -c_\alpha < x_k < c_\alpha) \geqq 1 - \alpha.$$

The proof of the following theorem establishes this inequality for certain cases

THEOREM. If $x_1, \cdots, x_k$ are normally distributed with zero means, unit variances, and correlations ρ_{is}, then

$$\int \underset{C}{\cdots} \int n_k(x_1, \cdots, x_k; 0, 1, \rho_{is})\, dx_1 \cdots dx_k \geqq \left[\int_{x=-c}^{x=c} n_1(x; 0, 1)\, dx\right]^k,$$

provided (1) $k = 2$ or 3; or (2) $\rho_{is} = b_i b_s$, for $i, s = 1, 2, \cdots, k, i \neq s$ and with $0 < b_i < 1, i = 1, 2, \cdots, k$. The region of integration C is the region bounded by the planes $x_i = \pm c, i = 1, \cdots, k$; $n_k(x_1, \cdots, x_k; 0, 1, \rho_{is})$ is the frequency function of $x_1, \cdots, x_k$; and $n_1(x; 0, 1)$ is the standard univariate normal frequency function.

PROOF. (1) $k = 2, 3$ For brevity the proof is merely outlined. The expression $\int \underset{C}{\cdots} \int n_k(x_1, \cdots, x_k; 0, 1, \rho_{is})\, dx_1, \cdots, dx_k$ may be regarded as a function of the ρ_{is}, say $F(\rho_{is})$. The proof consists in showing that for all admissible ρ_{is}, $F(\rho_{is})$ has an absolute minimum at the origin of the ρ_{is} space.

First it must be shown that there is a relative minimum at the origin This can be shown for any k by considering the various first and second partial derivatives with respect to the correlations.

The first partial derivative with respect to ρ_{12}, say F_{12}, can be shown to be:

$$F_{12} = 2 \int_{x_3=-c}^{x_3=c} \cdots \int_{x_k=-c}^{x_k=c} \cdot [n_k(c, c, x_3, \cdots, x_k; 0, 1, \rho_{is}) - n_k(c, -c, x_3, \cdots, x_k; 0, 1, \rho_{is})]\, dx_3, \cdots, dx_k \cdot$$

Similarly, the second derivative with respect to ρ_{12} and ρ_{pq}, say $F_{12,pq}$, is

$$F_{12,pq} = 2 \int_{x_3=-c}^{x_3=c} \cdots \int_{x_k=-c}^{x_k=c} n_k(c, c, x_3, \cdots, x_k; 0, 1, \rho_{is}) \cdot \left[\rho^{pq} + \left(\sum_{\substack{i=1 \\ x_1=c \\ x_2=c}}^{k} \rho^{pi} x_i\right)\left(\sum_{\substack{i=1 \\ x_1=c \\ x_2=c}}^{k} \rho^{qi} x_i\right)\right]$$

$dx_3, \cdots, dx_k$—a similar integral with $x_1 = c, x_2 = -c$.

When all the ρ_{is}'s are zero, it is easily seen that F_{12} vanishes. Further, $F_{12,pq}$ vanishes also at the origin unless $p = 1$ and $q = 2$, while $F_{12,12}$ is seen to be positive.

Thus in the expansion of $F(\rho_{is})$ about the origin, the first degree terms vanish and the second degree terms form a positive definite quadratic form, so that $F(\rho_{is})$ has a relative minimum at the origin for any k.

The next part of the proof is to show from the form of the first derivative, that at any point beside the origin, at least one of the first derivatives differ from zero. This was done only for $k = 2$ and 3.

The set of all admissible points (points such that (ρ_{is}) is positive definite and $0 < |(\rho_{is})| < 1$), together with the boundary points, form a compact set, so that $F(\rho_{is})$ must assume an absolute minimum either at an admissible point or at a boundary point. Hence if it can be shown that no point on the boundary of the set yields an absolute minimum, then the absolute minimum of F must be at the origin.

For $k = 2$, the boundary points are just $\rho_{12} = \pm 1$, and they actually yield absolute maxima for $F(\rho_{12})$.

For $k = 3$, a boundary point, say $(\rho_{12}, \rho_{13}, \rho_{23})$ was considered. It was shown that for m sufficiently close to 1 but less than 1, $(m\rho_{12}, m\rho_{13}, \rho_{23})$ is an admissible point, and that the derivative of F at $(m\rho_{12}, m\rho_{13}, \rho_{23})$ in the direction of $(\rho_{12}, \rho_{13}, \rho_{23})$ is positive. Hence $(\rho_{12}, \rho_{13}, \rho_{23})$ cannot yield an absolute minimum of F.

This completes the outline of the proof for $k = 2$ and 3, with any correlation matrix.

(2) For any k, if $\rho_{is} = b_i b_s$, with $0 < b_i < 1$ for $i = 1, \cdots, k$, a proof may be given which is adapted from the proof of a similar theorem by C. W. Dunnett and M. Sobel [3].

For $y_0, y_1, \cdots, y_k$ independently normally distributed, with zero means and unit variances, define

$$x_i = \sqrt{1 - b_i^2}\, y_i - b_i y_0, \qquad i = 1, \cdots, k.$$

Then the x_i's are normally distributed with means zero, unit variances, and correlations $\rho_{is} = b_i b_s$.

The theorem may be restated as follows:

$$P(-c < x_1 < c, \cdots, -c < x_k < c) \geqq \prod_{i=1}^{k} P(-c < x_i < c),$$

or

$$P(-c < \sqrt{1 - b_1^2}\, y_1 - b_1 y_0 < c, \cdots, -c < \sqrt{1 - b_k^2}\, y_k - b_k y_0 < c)$$
$$\geqq \prod_{i=1}^{k} P(-c < \sqrt{1 - b_i^2}\, y_i - b_i y_0 < c).$$

or

$$P(d_1 < y_1 < e_1, \cdots, d_k < y_k < e_k) \geqq \prod_{i=1}^{k} P(d_i < y_i < e_i),$$

where

$$d_i = \frac{-c + b_i y_0}{\sqrt{1 - b_i^2}}, \qquad e_i = \frac{c + b_i y_0}{\sqrt{1 - b_i^2}}, \qquad i = 1, \cdots, k.$$

This may be written as:

$$\int_{y_0=-\infty}^{y_0=\infty} \left[\int_{y_1=d_1}^{y_1=e_1} \cdots \int_{y_k=d_k}^{y_k=e_k} n_k(y_1, \cdots, y_k; 0, 1, 0)\, dy_1, \cdots, dy_k\right] n_1(y_0; 0, 1)\, dy_0$$
$$\geqq \prod_{i=1}^{k} \int_{y_0=-\infty}^{y_0=\infty} \left[\int_{y_i=d_i}^{y_i=e_i} n_1(y_i; 0, 1)\, dy_i\right] n_1(y_0; 0, 1)\, dy_0,$$

or

$$\int_{y_0=-\infty}^{y_0=\infty} \left[\prod_{i=1}^{k} F_i(y_0)\right] n_1(y_0; 0, 1)\, dy_0 \geqq \prod_{i=1}^{k} \int_{y_0=-\infty}^{y_0=\infty} F_i(y_0) n_1(y_0; 0, 1)\, dy_0,$$

where

$$F_i(y_0) = \int_{d_i}^{e_i} n_1(y_i; 0, 1)\, dy_i$$

Thus the inequality becomes:

$$E\left(\prod_{i=1}^{k} F_i(y_0)\right) \geqq \prod_{i=1}^{k} E(F_i(y_0))$$

The expected value of a product of monotone bounded functions is greater than or equal to the product of their expected values [6], so that the last inequality would hold if the F_i were monotone. The functions $F_i(y_0)$, however, are seen to increase from $-\infty$ to 0 and to decrease from 0 to ∞. Since the frequency function of y_0 is symmetric about the origin, the transformation $z = |y_0|$ changes the inequality to

$$E\left(\prod_{i=1}^{k} F_i(z)\right) \geqq \prod_{i=1}^{k} E(F_i(z)),$$

where $F_i(z)$ are monotonically decreasing bounded functions. This completes the proof of the theorem.

7.2. Variances unknown but equal. When the variances are unknown but equal, Student t-variables t_i with the joint frequency function

$$f_{n-1}(t_1, \cdots, t_k; \rho_{ij}),$$

as defined in 6.2, are used to form confidence intervals. Using the same methods as in 7.1, the following theorem can be proved:

THEOREM. *For* $k = 2$ *or* 3,

$$\int \underset{C}{\cdots} \int f_{n-1}(t_1, \cdots, t_k; \rho_{ij})\, dt_1 \cdots dt_k \geqq \int \underset{C}{\cdots} \int f_{n-1}(t_1, \cdots, t_k; 0)\, dt_1 \cdots dt_k$$

For $\rho_{is} = b_i b_s$, with $0 < b_i < 1$, $i = 1, \cdots, k$,

$$\int \underset{C}{\cdots} \int f_{n-1}(t_1, \cdots, t_k; \rho_{is})\, dt_1 \cdots dt_k \geqq \left[\int_{-c}^{c} f_{n-1}(t)\, dt\right]^k.$$

In this theorem, whose proof follows the same lines as the one in 7.1, C is the region bounded by $t_i = \pm c$, $i = 1, \cdots, k$, $f_{n-1}(t)$ is the density function of a Student t-variable with $n - 1$ degrees of freedom, and $f_{n-1}(t_1, \cdots, t_k; 0)$ is the joint frequency function of the t-variables when all ρ_{is} are zero.

Since

$$t_i = \frac{\sqrt{n}(\bar{y}_i - \mu_i)}{\sigma_1}, \qquad i = 1, \cdots, k,$$

sets of confidence intervals obtained are as follows:

For $k = 2$ or 3, $\bar{y}_1 \pm (\hat{\sigma}_1/n^{\frac{1}{2}})c_\alpha$, where c_α is defined by

$$\int_{-c_\alpha}^{c_\alpha} \cdots \int_{-c_\alpha}^{c_\alpha} f_{n-1}(t_1, \cdots, t_k; 0)\, dt_1 \cdots dt_k = 1 - \alpha.$$

For any k and $\rho_{is} = b_i b_s$, $0 < b_i < 1$, $i = 1, \cdots, k$, the same set is obtained, but with c_α defined by $H_{n-1}(c_\alpha) = (1 + (1 - \alpha)^{1/k})/2$, where H_{n-1} is the c.d.f. of a Student t-variable with $n - 1$ degrees of freedom. As in sections 5.2 and 6.2 one may use $\hat{\sigma}_p^2$ in place of $\hat{\sigma}_1^2$, provided one keeps $n - 1$ as the degrees of freedom.

8. Comparison of confidence intervals. In Table I are listed various sets of confidence intervals, with their properties and restrictions.

One rather obvious way to compare them is by comparing their lengths, or the expected values of the lengths. In Table II are given numerical values of d_α for $1 - \alpha = .95$, where

$$d_\alpha = \frac{\sqrt{n}}{\sigma} \sqrt{E(\tfrac{1}{2}\ell)^2},$$

with ℓ the length of the confidence interval. Throughout Table II, the variances are assumed to be equal.

When the variances are known and equal, and all the correlations are zero, the shortest set of confidence intervals must be those of section 7.1. When nothing is known about the correlations, no shorter set can be obtained. The last column in section 7 of Table II therefore gives the smallest obtainable values for d_α, and may be used as a standard for comparison.

For $1 - \alpha = .95$, the Bonferroni inequality intervals of section 6 are almost as good as the best ones. Indeed for $1 - \alpha$ as low as .80, the values of d_α are still very close, being:

k	Bonferroni	"Best"
1	1.28	1.28
2	1.64	1.61
4	1.96	1.92
6	2.13	2.09
8	2.24	2.20
10	2.33	2.29

TABLE I

Confidence Intervals for Means of Dependent, Normally Distributed Variables

Section	Confidence Intervals	Definition of c_α	Conditions
3 1	$\sum_{j=1}^{n} a_{ij} y_{ij} \pm \frac{k}{n} \sigma_i \cdot c_\alpha$	$N(c_\alpha) = \frac{1 + (1-\alpha)^{1/k}}{2}$	$n \geq k$ (1)
3 2	$\sum_{j=1}^{n} a_{ij} y_{ij} \pm \sqrt{\frac{k}{n(n-k)} \sum_{m=1}^{n-k} u_m^2}\; c_\alpha$	$H_{n-k}(c_\alpha) = \frac{1 + (1-\alpha)^{1/k}}{2}$	$n > k$ (2, 3)
4.1	$\bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} \cdot \sqrt{c_\alpha}$	$U_k(c_\alpha) = 1 - \alpha$	(4)
4 2	$\bar{y}_i \pm \frac{\hat{\sigma}_i}{\sqrt{n}} \cdot c_\alpha$	$F(c_\alpha) = 1 - \alpha$	$n > k$ (5)
5 1	$\bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} \sqrt{c_\alpha}$	$U_k(c_\alpha) = 1 - \alpha$	(4)
5 2	$\bar{y}_i \pm \frac{\hat{\sigma}_1}{\sqrt{n}} \sqrt{c_\alpha}$	$F_{k,n-1}(c_\alpha) = 1 - \alpha$	(6)
6 1	$\bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} \cdot c_\alpha$	$N(c_\alpha) = 1 - \frac{\alpha}{2k}$	(1)
6 2	$\bar{y}_i \pm \frac{\hat{\sigma}_1}{\sqrt{n}} \cdot c_\alpha$	$H_{n-1}(c_\alpha) = 1 - \frac{\alpha}{2k}$	(2)
7 1	$\bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} \cdot c_\alpha$	$N(c_\alpha) = \frac{1 + (1-\alpha)^{1/k}}{2}$ $k = 2, 3,$ or $\rho_{is} = b_i b_s$	(1)
7 2	$\bar{y}_i \pm \frac{\hat{\sigma}_1}{\sqrt{n}} \cdot c_\alpha$	$H_{n-1}(c_\alpha) = \frac{1 + (1-\alpha)^{1/k}}{2}$ $k = 2, 3,$ or $\rho_{is} = b_i b_s$	 (2, 3) (2)

(1) N is the cumulative standard normal distribution function

(2) H_ν is the cumulative distribution function of a Student t-variable with ν degrees of freedom.

(3) This definition of c_α is approximate. The exact definition is

$$\int_{-c_\alpha}^{c_\alpha} \cdots \int_{-c_\alpha}^{c_\alpha} f_\nu(t_1, \cdots, t_k)\, dt_1, \cdots, dt_k = 1 - \alpha, \quad \text{where } f_\nu(t_1, \cdots, t_k)$$

$$= \frac{\Gamma\left(\frac{k+\nu}{2}\right)}{\nu^{k/2}\pi^{k/2}\Gamma\left(\frac{\nu}{2}\right)} \left[1 + \frac{\sum_{i=1}^{k} t_i^2}{\nu}\right]^{-(k+\nu)/2}, \quad \text{where}$$

ν is the degrees of freedom of t_i.

(4) U_k is the cumulative distribution function of a Chi square variable with k degrees of freedom.

(5) F is the cumulative distribution function of Hotelling's T.

(6) $F_{k,n-1}$ is the cumulative distribution function of an F variab[illegible] n degrees of freedom

TABLE II

*Comparison of Lengths of Confidence Intervals for Means of Dependent, Normally Distributed Variables with Equal Variances, $1 - \alpha = .95$**

k	n							
	Variances Unknown						Variances Known	
	Section	4	6	8	10	20	Section	Any
	4.1						4.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		10.8	5.52	4.53	4.12	3.55		3.17
4			27.9	11.4	8.63	6.13		4.98
6				49.3	17.8	8.75		6.44
8					77.0	11.8		7.72
10						15.6		8.85
	5.1						5.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		7.55	4.16	3.47	3.17	2.74		2.45
4			13.9	6.70	5.21	3.79		3.08
6				20.1	9.12	4.82		3.55
8					26.4	6.01		3.94
10						7.53		4.28
	6.1						6.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		4.37	3.40	3.08	2.92	2.66		2.45
4		6.04	4.56	4.06	3.81	3.41		3.08
6		7.32	5.45	4.82	4.50	3.98		3.55
8		8.41	6.21	5.37	5.08	4.46		3.94
10		9.38	6.88	6.03	5.60	4.89		4.28
	7.1						7.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		4.17	3.16	2.84	2.68	2.44		2.24
4		5.41	3.80	3.33	3.11	2.76		2.50
6		6.22	4.22	3.64	3.36	2.94		2.64
8		6.92	4.53	3.86	3.55	3.07		2.74
10		7.47	4.77	4.03	3.69	3.17		2.81
	8.1						8.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		4.16	3.15	2.83	2.68	2.43		2.24
4		5.35	3.79	3.32	3.10	2.75		2.49
6		6.17	4.20	3.62	3.35	2.94		2.63
8		6.86	4.50	3.84	3.53	3.07		2.73
10		7.40	4.76	4.01	3.67	3.16		2.80

* The figures given in the table are values of $(n^{\frac{1}{2}}/\sigma)\sqrt{E(\frac{1}{2}\ell)^2}$, where ℓ is the length of the confidence interval.

It would be interesting to show that the "best" intervals can be used for arbitrary k and arbitrary correlations, but from a practical viewpoint, for $1 - \alpha$ large enough to be of interest, the Bonferroni regions are good enough.

The regions of section 5, based on the T-distribution and the χ^2 distribution, compare favorably only when k is small and n relatively large. The regions with exact confidence level are everywhere unnecessarily long

REFERENCES

[1] PAUL G. HOEL, "Confidence bands for polynomial curves," *Ann Math Stat*, Vol 25, (1954), pp. 534-542.

[2] C. W. DUNNET AND M. SOBEL, "Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution," *Biometrika*, Vol. 42, (1955), pp 258-260

[3] H. B. MANN AND A. WALD, "On stochastic limit and order relationships," *Ann Math Stat*, Vol 14, (1943), p. 224.

[4] HENRY SCHEFFÉ, "A method of judging all contrasts in the analysis of variance," *Biometrika*, Vol. 40, (1953), pp 87-104.

[5] W. FELLER, *An introduction to probability theory and its applications* I, p 75, John Wiley and Sons, 1950.

[6] E. PAULSON, "On the comparison of several experimental categories with a control," *Ann. Math. Stat.*, Vol. 23, (1952), p. 249.

[7] A W. KIMBALL, "On dependent tests of significance in the analysis of variance," *Ann. Math. Stat.*, Vol 22, (1951), pp 600-602

TABLE II

*Comparison of Lengths of Confidence Intervals for Means of Dependent, Normally Distributed Variables with Equal Variances, $1 - \alpha = .95$**

k	Variances Unknown: Section	n = 4	n = 6	n = 8	n = 10	n = 20	Variances Known: Section	Any
	4.1						4.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		10.8	5.52	4.53	4.12	3.55		3.17
4			27.9	11.4	8.63	6.13		4.98
6				49.3	17.8	8.75		6.44
8					77.0	11.8		7.72
10						15.6		8.85
	5.1						5.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		7.55	4.16	3.47	3.17	2.74		2.45
4			13.9	6.70	5.21	3.79		3.08
6				20.1	9.12	4.82		3.5[illegible]
8					26.4	6.01		3.[illegible]
10						7.53		
	6.1						6.2	
1		3.18	2.57	2.36	2.26	2.09		
2		4.37	3.40	3.08	2.92	2.66		
4		6.04	4.56	4.06	3.81	3.41		
6		7.32	5.45	4.82	4.50	3.98		
8		8.41	6.21	5.37	5.08	4.46		
10		9.38	6.88	6.03	5.60	4.89		
	7.1							
1		3.18	2.57	2.36	2.26	2.0[illegible]		
2		4.17	3.16	2.84	2.68	2.[illegible]		
4		5.41	3.80	3.33	3.11	[illegible]		
6		6.22	4.22	3.64	3.36			
8		6.92	4.53	3.86	3.55			
10		7.47	4.77	4.03	3.69			
	8.1							
1		3.18	2.57	2.36	2.[illegible]			
2		4.16	3.15	2.83	2.[illegible]			
4		5.35	3.79	3.32	[illegible]			
6		6.17	4.20	3.62				
8		6.86	4.50	3.84				
10		7.40	4.76	4.01				

* The figures given in the table are values of [illegible]
confidence interval.

continuous in t. Assume that

$$\lim_{t \downarrow 0} P(t) = I.$$

Clearly

$$P(t)P(s) = P(t + s), \qquad t, s > 0.$$

Let the initial probability distribution of $X(t)$ be w, $w_i > 0$, $i = 1, \cdots, m$. Then $Y(t) = f(X(t))$ is Markovian, whatever the initial distribution w of $X(t)$, if and only if for each $\beta = 1, \cdots, r$ separately either

$$\text{(7)} \quad \begin{aligned} &\text{(i)} \quad p_{i,s_\beta}(t) = 0 \quad \text{for all} \quad i \varepsilon S_\beta \quad \text{or} \\ &\text{(ii)} \quad p_{i,s_\gamma}(t) = C_{s_\beta,s_\gamma}(t) \quad \text{for every} \quad i \varepsilon S_\beta \quad \text{and all} \quad \gamma = 1, \cdots, r. \end{aligned}$$

Part of the interest in the proofs of Theorems 1 and 4 lies in the fact that they show that if the collapsed processes in these cases satisfy the Chapman-Kolmogorov equations, they are Markovian.

Condition (3) can be reworded in the case of a Markov process $X(t)$, $0 \leq t < \infty$, with stationary transition probabilities and values in an abstract space. Let Ω be a space of points x and $B(\Omega)$ a Borel field on Ω. Further let the sets (x) be elements of $B(\Omega)$. Consider a function

$$P(t; x, A), \qquad A \varepsilon B(\Omega)$$

satisfying

(i) $P(t; x, A)$ is a Baire function of x for fixed t, A;
(ii) $P(t; x, A)$ is a probability measure in $A \varepsilon B(\Omega)$ for fixed t, x;
(iii) $P(t; x, A)$ satisfies the Chapman-Kolmogorov equation

$$P(t + \tau; x, A) = \int_\Omega P(t; y, A)P(\tau; x, dy), \qquad t, \tau > 0.$$

Let $X(t)$ be a Markov chain with $P(t; x, A)$ as its transition probability function. Let f be a function from Ω onto another space of points Ω'. The function f induces a Borel field of sets $B(\Omega') = f(B(\Omega))$ on Ω'. This consists of sets of the form $fA = (y \varepsilon \Omega' \mid y = f(x), x \varepsilon A)$, $A \varepsilon B(\Omega)$. Now consider the inverse images of sets in $f(B(\Omega))$. The class of sets of this form we call $f^{-1}f(B(\Omega))$ and it is a subBorel field of $B(\Omega)$ consisting of sets of the form

$$\{z \varepsilon \Omega \mid z = f^{-1}f(x), x \varepsilon A\}, \qquad A \varepsilon B(\Omega).$$

The analogue of condition (3) is simply that

$$\text{(8)} \qquad P(t; x, A), \qquad A \varepsilon f^{-1}f(B(\Omega))$$

be a Baire function of x with respect to $f^{-1}f(B(\Omega))$ for fixed t, A.

Corollary 3. *$Y(t) = f(X(t))$ is a Markov process, whatever the initial probability distribution of $X(t)$, if condition* (8) *is satisfied.* Condition (8) is discussed

in a paper of B. Rankin [4] as a sufficient condition for a collapsed Markovian process to be Markovian.

2. The stationary case. Let the assumptions of Theorem 1 be satisfied. The matrix of n-step transition probabilities of the process $Y(n)$ is of the form

$$(9) \qquad Q^{(n)} = AP^nB = (q_{\alpha\beta}^{(n)}) = (P[X(t+n) \in S_\beta \mid X(t) \in S_\alpha]),$$

where A, B are $r \times m$ and $m \times r$ matrices respectively The elements of B are of the form

$$b_{ij} = \begin{cases} 1 & \text{if } \quad i \in S_j, \\ 0 & \text{otherwise}; \end{cases}$$

while

$$(10) \qquad A = (B'DB)^{-1}B'D,$$

where D is the diagonal matrix introduced above If the new process is Markovian, the Chapman-Kolmogonov equation must be satisfied by the $Q^{(n)}$, that is,

$$(11) \qquad Q^{(n)} = AP^nB = [Q^{(1)}]^n = (APB)^n, \qquad n = 2, 3, \cdots$$

This condition can be reworded in an equivalent form

$$(12) \qquad AP^nBAPB = AP^{n+1}B, \qquad n = 1, 2, 3, \cdots$$

Note that

$$(13) \qquad BAPB = PB$$

implies that (12) is satisfied. Condition (13) is just condition (3) expressed in matrix form when the assumptions of Theorem 1 are satisfied. We first verify that (3) implies that $Y(n)$ is Markovian (To facilitate printing we sometimes write $\alpha(i)$ in place of α_i.) Clearly

$$P[Y(0) \in S_{\alpha(0)}, \cdots, Y(n) \in S_{\alpha(n)}] = \sum_{j=0}^{n} \sum_{i_j \in S_{\alpha(j)}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$$
$$= \Big(\sum_{i \in S_{\alpha(0)}} p_i \Big) C_{S_{\alpha(0)}, S_{\alpha(1)}} \cdots C_{S_{\alpha(n-1)}, S_{\alpha(n)}}$$

and it is easily seen that

$$C_{S_\alpha, S_\beta} = P[Y(n+1) \in S_\beta \mid Y(n) \in S_\alpha].$$

The sufficiency of condition (3) is thus verified. Note that the sufficiency argument given above holds for the case of any initial distribution ϖ and without the condition of reversibility. We thus have Corollary 1.

Let us now consider the necessity of condition (3) when $X(n)$ is reversible. If $Y(n)$ is Markovian the Chapman-Kolmogorov equations are satisfied by the $Q^{(n)}$ and we must have

$$Q^{(2)} = [Q^{(1)}]^2$$

or

$$AP(I - BA)PB = 0.$$

But this implies that

$$B'DP(I - BA)PB = 0.$$

Because of reversibility, this can be written as

$$B'P'D(I - BA)PB = 0.$$

Now $D(I - BA)$ is positive definite so that

$$D(I - BA) = R'R$$

for some $m \times m$ matrix R. Thus

$$(RPB)'(RPB) = 0$$

and

$$RPB = 0.$$

But then

$$R'RPB = D(I - BA)PB = 0$$

and hence

$$(I - BA)PB = 0.$$

It is worth while noting that the problems we consider are related to issues of aggregation and consolidation in multisector models of mathematical economics (see [5]). There one has a stochastic matrix P and an invariant vector

$$p, pP = p.$$

One asks for the types of aggregation under which the aggregated invariant vector is an invariant vector of the aggregated matrix. The aggregated matrix $Q = APB$ where B is defined as before and $A = (B'D_vB)^{-1}B'D_v$. Here D_v is the diagonal matrix with its ith diagonal element v_i. The aggregation is determined by the sets of states S_i and the vector $v = (v_i)$. The aggregated vector is pB. The question is then for what aggregation schemes the relation

$$pBQ = pB(B'D_vB)^{-1}B'D_vPB = pB$$

is valid. Conditions (3) and (6) turn out to be crucial in some of the results obtained in [5].

3. Any initial distribution. Let the assumptions of Theorem 2 be satisfied. We first show that (4) is sufficient. It is enough to show that

$$\begin{aligned} P[X(n) = i, X(n+1) \varepsilon S, \cdots, X(n+h) \varepsilon S, X(n+h+1) = j] \\ = P[X(n) = i]P[X(n+1) \varepsilon S \mid X(n) = i] \\ \cdots P[X(n+h) \varepsilon S \mid X(n+h-1) \varepsilon S] \\ P[X(n+h+1) = j \mid X(n+h) \varepsilon S] \end{aligned}$$

for any $j \varepsilon S$ and any i, since then $Y(n)$ is clearly Markovian. Note that (1) implies that

$$(14) \qquad \sum_{l \varepsilon S} p_{kl}\, p_{l,S} = p_{k,S}\, C_S$$

for all k. By making use of (4) and (14) the following relation is obtained

$$P[X(n+h+1) = j, X(n+h) \varepsilon S, \cdots, X(n+1) \varepsilon S \mid X(n) = i]$$
$$= \sum_{k=1}^{h} \sum_{i_k \varepsilon S} p_{i\,i_1}\, p_{i_1,i_2} \cdots p_{i_{k-1},i_k}\, p_{i_k j}$$
$$= p_{i,S}(C_S)^{h-1} C_j .$$

But

$$C_j = P[X(n+1) = j \mid X(n) \varepsilon S], \qquad j \varepsilon S,$$

and

$$C_S = P\{X(n+1) \varepsilon S \mid X(n) \varepsilon S\}.$$

An Argument paralleling the one given above indicates that (1') implies that $Y(n)$ is Markovian so that we have Corollary 2 $Y(n)$ is obviously Markovian if (5) is satisfied.

Now consider the necessity of (1). Since $Y(n)$ is Markovian whatever the initial distribution w of $X(n)$, the transition probabilities of $Y(n)$ satisfy the Chapman-Kolmogorov equation. It may be that $p_{i,S} = 0$ for all i. Then (1) is obviously satisfied. Suppose now that there is an i such that $p_{i,S} \neq 0$. The Chapman-Kolmogorov equation then tells us that

$$p_{i,S} \frac{\sum_{l \varepsilon S} \sum_{k} w_k\, p_{kl}\, p_{lu}}{\sum_{k} w_k\, p_{kS}} = \sum_{l \varepsilon S} p_{il}\, p_{lu}$$

for all i, $u \varepsilon S$. If k is such that $p_{k,S} \neq 0$ then

$$(15) \qquad p_{i,S} \sum_{l \varepsilon S} p_{kl}\, p_{lu} = p_{k,S} \sum_{l \varepsilon S} p_{il}\, p_{lu}$$

as is seen by letting $w_k \to 1$ and $w_l \to 0$, $l \neq k$. And if $p_{k,S} = 0$ (15) is obviously satisfied. Thus (15) holds for all k and all $i \varepsilon S$. If there is an $i \varepsilon S$ such that $p_{i,S} \neq 0$ (15) is satisfied for all k and i. But this implies relation (1). There is still the possibility that $p_{i,S} = 0$ for all $i \varepsilon S$, namely condition (5).

In the context of Theorem 2 condition (3) implies that condition (1) is satisfied. However, the converse is not true. Consider the transition probability matrix

$$P = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$

Collapse the states 1, 2, 3 into a set S and leave the states 4, 5 alone. Note that (3) is not satisfied. But (4) is satisfied since

$$\frac{\sum_{l \varepsilon S} p_{kl}\, p_{lu}}{p_{k,S}} = \tfrac{1}{6}$$

for all $u \varepsilon S$ and all k.

4. Any function f. The answer obtained to the question posed in Theorem 3 is the same as the answer obtained in a similar problem posed by Bush, Mosteller and others [1]. The structure of interest in Bush and Mosteller's problem is not Markovian. Note that in our case we ask that $f(X(n))$ have the same structure (a Markovian structure) as $X(n)$ for any f and a specific initial probability vector, a left invariant vector p of P. Bush and Mosteller ask that $f(X(n))$ have the same structure as $X(n)$ for any f and any initial probability vector w.

Let us now prove Theorem 3. The condition imposed on the process will not be used in full strength. Just consider a consolidation in which two states j, k are consolidated into a set S and all other states are left the same. Let i, l be any indices distinct from j, k. Since the consolidated process is Markovian, its transition probabilities satisfy the Chapman-Kolmogorov equation and hence

$$(16) \qquad p_{il}^{(2)} = \sum_{u=1}^{m} p_{iu}\, p_{ul} = \sum_{u \varepsilon S} p_{iu}\, p_{ul} + (p_{ij} + p_{ik}) \frac{p_j\, p_{jl} + p_k\, p_{kl}}{p_j + p_k}.$$

Equation (16) can be reduced to the following convenient form

$$(17) \qquad (p_{ij}p_k - p_{ik}p_j)(p_{jl} - p_{kl}) = 0.$$

Further, (17) implies that

$$(18) \qquad [(p_jp_{jj} + p_kp_{kj})p_k - (p_jp_{jk} + p_kp_{kk})p_j](p_{jl} - p_{kl}) = 0.$$

First consider the case in which for all i $p_{ij}p_k = p_{ik}p_j$ for all $j, k \neq i$. But then

$$p_{ij} = (1 - \lambda_i)p_j, \qquad i \neq j,$$

$$\lambda_i = \frac{p_{ii} - p_i}{1 - p_i},$$

so that P is of the form

$$P = \Lambda + (I - \Lambda)U,$$

where Λ is a diagonal matrix with diagonal elements λ_i and U is a matrix with identical rows $(p_1, \cdots, p_n)$. If

$$(19) \qquad (p_jp_{jj} + p_kp_{kj})p_k = (p_jp_{jk} + p_kp_{kk})p_j$$

for some pair of indices j, k it follows that $\lambda_j = \lambda_k$. If (19) does not hold for the pair j, k, (18) implies that $p_{jl} = p_{kl}$ for all $l \neq j, k$. But then $\lambda_j = \lambda_k$. Thus it follows that in this case $\lambda_1 = \lambda_2 = \cdots = \lambda_n$.

Now on the contrary assume there is a row i for which $p_{ij}p_k = p_{ik}p_j$ does not hold for all $j, k \neq i$. Given any $j \neq i$ consider all k for which we can find a sequence $j_1, \cdots, j_\alpha$ such that

$$p_{ij}p_{j_1} = p_{ij_1}p_j, \quad p_{ij_1}p_{j_2} = p_{ij_2}p_{j_1}, \cdots, p_{ij_\alpha}p_k = p_{ik}p_{j_\alpha}.$$

There is a maximal set of such indices k (including j of course). There are at least two such sets. The collection of all such maximal sets are disjoint. Given any j in one such maximal set and any j' in another we must have

$$(20) \qquad p_{jl} = p_{j'l}$$

for all $l \neq j, j'$ and

$$(21) \qquad p_{jj'} + p_{jj} - p_{j'j} - p_{j'j'} = 0.$$

For convenience let us assume $i = 1$. Keeping (20) and (21) in mind, it is clear that for any fixed $j \neq 1$ the p_{kj}'s must be equal for all $k \neq 1, j$. Call this common value u_j. Thus all rows except possibly for the first must be of the form

$$p_{kj} = \lambda\delta_{kj} + u_j.$$

There are now two possibilities. Either $p_{ij}p_k = p_{ik}p_j$ for all $i \neq 1$ and all

$$j, k \neq i$$

or this is not the case. If not we must have $p_{ij} = \lambda\delta_{ij} + u_j$ for all i. Since p is an invariant vector $u_j = (1 - \lambda)p_j$. On the other hand if $p_{ij}p_k - p_{ik}p_j = 0$ for all $i \neq 1$ and $j, k \neq i$ then $u_j = (1 - \lambda)p_j$. The elements of the first row are as yet unknown. But again making use of the fact that p is a stationary distribution we see that $p_{ij} = \lambda\delta_{1j} + (1 - \lambda)p_j$.

5. Finite state space and continuous time. The proof of the sufficiency of condition (7) in the case of Theorem 4 parallels the proof of Corollary 1.

We now show that (7) is necessary. A transition probability matrix-valued function $P(t)$ satisfying the regularity conditions posed in the assumptions in Theorem 4 is of the form (see [2])

$$P(t) = \exp(Gt),$$

where $G = (g_{ij})$ is such that

$$g_{ij} \geqq 0, \qquad i \neq j,$$

$$\sum_{\substack{j=1 \\ j \neq i}}^{m} g_{ij} = -g_{ii}.$$

Let $w = (w_i)$, $w_i > 0$ be the initial distribution of $X(t)$. A necessary condition that the collapsed process be Markovian for an initial vector can be written down conveniently in matrix notation. As before, let

$$Q_w^{(t)} = (B'D_wB)^{-1}B'D_wP(t)B$$

denote the t-step transition probability matrix (from time zero to time t) for the collapsed process $Y(t)$ when the initial probability distribution vector of the original process $X(t)$ is w. If the collapsed process $Y(t)$ is Markovian $Q_w^{(t)}$ must satisfy the Chapman-Kolmogorov equation and thus

$$Q_w^{(t)} Q_{wP(t)}^{(\tau)} = Q_w^{(t+\tau)}, \qquad t, \tau > 0, \tag{22}$$

for all w, $w_i > 0$. It is clear that the w_i's only have to satisfy $w_i > 0$ and that the condition $\sum w_i = 1$ needn't be imposed. On differentiating relationship (22) with respect to τ at $\tau = 0$ we obtain

$$Q_w^{(t)}(B'D_{wP(t)}B)^{-1}B'D_{wP(t)}GB = (B'D_wB)^{-1}B'D_wP(t)GB. \tag{23}$$

Let us now differentiate (23) with respect to t at $t = 0$. We then have

$$B'D_wGB(B'D_wB)^{-1}B'D_wGB - (B'D_wB)^{-1}B'D_{wG}BB'D_wGB + B'D_{wG}GB = B'D_wG^2B.$$

This can be written more conveniently as

$$B'[D_wG - G_{wG}][B(B'D_wB)^{-1}(B'D_w) - I]GB = 0. \tag{24}$$

Let

$$w_{S_\beta} = \sum_{i \varepsilon S_\alpha} w_i\,,$$

$$g_{i,S_\beta} = \sum_{j \varepsilon S_\alpha} g_{ij}\,.$$

Condition (24) can be written down elementwise as

$$\sum_{i \varepsilon S_\alpha} \sum_{\gamma} w_i\, g_{i,S_\gamma}\, w_{S_\gamma}^{-1} \sum_{i \varepsilon S_\gamma} w_i\, g_{i,S_\beta} - \sum_{i \varepsilon S_\alpha} \sum_{k} w_i\, g_{ik}\, g_{k,S_\beta}$$
$$- \sum_{i} w_i\, g_{i,S_\alpha}\, w_{S_\alpha}^{-1} \sum_{i \varepsilon S_\alpha} w_i\, g_{i,S_\beta} + \sum_{i} w_i \sum_{k \varepsilon S_\alpha} g_{ik}\, g_{k,S_\beta} = 0. \tag{25}$$

If we set $w_i = u_i h$, $i \varepsilon S_\alpha$, in (25) and then let $h \downarrow 0$, the following relation is obtained since the first two terms drop out

$$-\sum_{i \notin S_\alpha} w_i\, g_{i,S_\alpha}\, u_{S_\alpha}^{-1} \sum_{i \varepsilon S_\alpha} u_i\, g_{i,S_\beta} + \sum_{i \notin S_\alpha} w_i \sum_{k \varepsilon S_\alpha} g_{i,k}\, g_{k,S_\beta} = 0.$$

But this is valid if and only if

$$g_{i,S_\alpha}\, u_{S_\alpha}^{-1} \sum_{i \varepsilon S_\alpha} u_i\, g_{i,S_\beta} = \sum_{k \varepsilon S_\alpha} g_{i,k}\, g_{k,S_\beta}$$

for all $i \notin S_\alpha$. Further, since this holds for all u_i,

$$g_{i,S_\alpha}\, g_{j,S_\beta} = \sum_{k \varepsilon S_\alpha} g_{i,k}\, g_{k,S_\beta} \tag{26}$$

for all $i \notin S_\alpha$ and all $j \varepsilon S_\alpha$. There are only two alternatives that arise. If

$$g_{i,S_\alpha} = 0$$

for all $i \varepsilon S_\alpha$ relationship (26) is obviously satisfied (we then say that S_α satisfies (i)). Otherwise $g_{i,S_\alpha} \neq 0$ for some $i \varepsilon S_\alpha$ in which case g_{j,S_β} for each β is a constant for all $j \varepsilon S_\alpha$, that is,

$$g_{j,S_\beta} = K_{S_\alpha,S_\beta} \tag{27}$$

for all $j \varepsilon S_\alpha$, $\beta = 1, \cdots, r$ (we then say that S_α satisfies (ii)). The matrix G is said to satisfy (7) if for each α separately S_α satisfies either (i) or (ii). Note that if G satisfies (7) the nth power of G, $G^n = (g_{ij}^{(n)})$, satisfies (7) in a consistent manner, that is, S_α satisfies (i) for G^n if and only if S_α satisfies (i) for G. Since

$$P(t) = \exp(Gt) = \sum_{k=0}^{\infty} G^k t^k / k!$$

$P(t)$ satisfies (7). It should be noted that our proof has shown that the condition that the Chapman-Kolmogorov equation be satisfied by the collapsed process is enough to imply that the new process be Markovian. P. Levy [3] has shown that this is generally not the case.

6. Abstract state space. Consider a Markov process $X(t)$ with initial probability distribution

$$P[X(0) \varepsilon A] = P(A), \qquad A \varepsilon B(\Omega)$$

and transition probability function

$$P(t, x, A)$$

satisfying the assumptions of Corollary 3. Then $Y(t) = f(X(t))$ is a Markovian process with initial distribution

$$P[Y(0) \varepsilon A'] = P[X(0) \varepsilon f^{-1}(A')] = Q(A')$$

$A' \varepsilon f(B(\Omega))$, and transition probability function

$$\begin{aligned} Q(t; y, A') &= P[Y(t + \tau) \varepsilon A' \mid Y(\tau) = y] \\ &= P[X(t + \tau) \varepsilon f^{-1}(A') \mid X(\tau) \varepsilon f^{-1}(y)] \\ &= P(t; x, f^{-1}(A')), \; y \varepsilon \Omega', \, A' \varepsilon f(B(\Omega)), \end{aligned}$$

where x is such that $y = f(x)$. This follows immediately from condition (8).

It is interesting to note that one can generate new Markovian processes from old ones by setting up f so that it is consistent with the symmetries of the transition probability mechanism of the old process. Consider $X(t)$ Brownian motion on the line. Here the transition probability density is

$$P(t; x, y) = (2\pi t)^{-1/2} \exp\left(-\frac{1}{2t}(x - y)^2\right), \qquad t > 0.$$

If we set

$$f(x) = x - a[x/a], \qquad a > 0,$$

where $[x]$ is the greatest integer less than or equal to x, the new Markovian process $Y(t) = f(X(t))$ is Brownian motion on the circle. If

$$f(x) = z$$

on all points of the form $2ka \pm z$, $0 \leqq z < a$, $k = 0, \pm 1, \cdots$, $Y(t)$ is Brownian motion on a line segment of length a with reflecting barriers at the endpoints.

As a further example consider starting out with two-dimensional Brownian motion $(X_1(t), X_2(t))$, that is, the transition probability density is

$$p(t; (x_1, x_2), (y_1, y_2)) = (2\pi t)^{-1} \exp\left(-\frac{1}{2t}\left[\left(x_1 - y_1\right)^2 + \left(x_2 - y_2\right)^2\right]\right), \quad t > 0.$$

If

$$f(x_1, x_2) = (u_1, u_2)$$

for all points (x_1, x_2) of the form $(u_1 + ja, u_2 + ka)$ $0 \leqq u_1, u_2 < a$, $j, k = 0, \pm 1, \cdots$ $(Y_1(t), Y_2(t))$ is Brownian motion on a torus. If

$$f(x_1, x_2) = (u_1, u_2)$$

for all points of the form $(u_1 + ja, (2k + j)\, a \pm u_2)$ $0 \leqq u_1, u_2 < a$, $j, k = 0, \pm 1, \cdots$ $(Y_1(t), Y_2(t))$ is Brownian motion on a Moebius strip with reflecting barriers on the edges of the strip.

REFERENCES

[1] BUSH AND MOSTELLER, *Stochastic Models for Learning*, John Wiley & Sons, New York.
[2] J. L. DOOB, *Stochastic Processes*, John Wiley & Sons, 1953.
[3] P. LEVY, "Examples de processus pseudo-Markoviens," *C. R. Acad. Sci. Paris*, Vol. 228 (1949), p. 2004–2006.
[4] B. RANKIN, "The concept of enchainment—a relation between stochastic processes" (1955), unpublished.
[5] D. ROSENBLATT, "On aggregation and consolidation in linear systems", to be published in the *Naval Research Logistics Quarterly*.

ASYMPTOTIC DISTRIBUTIONS OF "PSI-SQUARED" GOODNESS OF FIT CRITERIA FOR m-TH ORDER MARKOV CHAINS[1]

By Leo A. Goodman

University of Chicago

1. Introduction and Summary. Let $\{X_1, X_2, \cdots, X_N\}$ be an observed sequence from a stochastic process, where X_i can take any one of s values 1, 2, $\cdots$, s. Let f_u be the frequency of the m-tuple $u = (u_1, u_2, \cdots, u_m)$ in the sequence. Let H'_n be the composite hypothesis that the process is a Markov chain of order n. Let H_n be any simple hypothesis belonging to H'_n. Let H^*_n be the maximum likelihood H_n. Let the expected value of f_u in a new sequence of length N given H_n be $f_{u,n}$, and given H^*_n be $f^*_{u,n}$. Let

$$\psi^2_{m,n} = \sum_u (f_u - f_{u,n})^2/f_{u,n},$$

$$\psi^{*2}_{m,n} = \sum_u (f_u - f^*_{u,n})^2/f^*_{u,n},$$

$$\psi^{*2}_{n+1,n} = 0 .$$

Good had proposed in [7] the following two conjectures: (a) that the asymptotic distribution ($N \to \infty$) of $\psi^{*2}_{m,n}$, when H'_n is true, is

$$\mathop{*}_{\lambda=1}^{m-n-1} K_{g(\lambda)}\,(x/\lambda) ,$$

where $*$ denotes convolution, $g(\lambda) = (s-1)^2 s^{m-1-\lambda}$, and $K_i(x)$ is the χ^2-distribution with i degrees of freedom, (b) that the asymptotic distribution of $\psi^2_{m,n}$, when H_n is true, is

$$\mathop{*}_{\lambda=1}^{m-1} K_{g(\lambda)}(x/\lambda) * K_{s-1}(x/m) ,$$

mathematically independent of n. Conjectures (a) and (b) were proved by Billingsley [2] for the special case $n = 0$. For the special case $n = -1$ (by convention, H'_{-1} is the hypothesis of equiprobable or perfect randomness (see [7])), Conjecture (b) was proved by Good [5] when s is prime. In the present paper, Conjecture (a) will be proved for the general case $n \geqq -1$; conjecture (b) will be shown to be incorrect for $n > 0$, although a modified version of (b) will be proved for $n \geqq -1$. A third conjecture by Good [6] will also be proved here. It was

Received December 13, 1957.

[1] Research carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

I am indebted to I. J. Good for the opportunity to read [7] before its publication, and also for mentioning that he suspected that his conjectures in [7] could be proved with the aid of the results in my earlier paper [10].

assumed in these earlier papers, and it will be assumed here, that all transition probabilities in the Markov chain are positive; the results can be modified accordingly when some of these probabilities are zero (see [1] and [10]).

Let $M_{m,n} = -2 \log \lambda_{n,m-1}$, where $\lambda_{n,m-1}$ is the ratio of the maximum likelihood given H'_n to that given H'_{m-1} (see [6]). For $m = n + 2$, the statistic $\psi^{*2}_{m,n}$ is asymptotically equivalent, when H'_n is true, to the likelihood ratio statistic $M_{m,n}$. For $m > n + 2$, $\psi^{*2}_{m,n}$ is asymptotically equivalent, when H'_n is true, to $\sum_{\lambda=1}^{m-n-1} \lambda M_{m+1-\lambda,m-1-\lambda}$, while $M_{m,n}$ is asymptotically equivalent to

$$\sum_{\lambda=1}^{m-n-1} M_{m+1-\lambda,m-1-\lambda}$$

(see [6], [10]). Thus, $\psi^{*2}_{m,n}$ corresponds asymptotically to a weighted sum of the likelihood ratio statistics $M_{n+2,n}, M_{n+3,n+1}, \cdots, M_{m,m-2}$, with the weights $m - n - 1, m - n - 2, \cdots, 1$, respectively, while $M_{m,n}$ weights these statistics equally (see [13] and reference to [13] in Section 4 herein).

Let $L_{m,n} = -2 \log \mu_{n,m-1}$, where $\mu_{n,m-1}$ is the ratio of the likelihood given H_n to the maximum likelihood given H'_{m-1}. For $m - 1 = n = 0$, the statistic $\psi^2_{m,n}$ is asymptotically equivalent, when H_n is true, to $L_{m,n}$. For $m - 1 > n = 0$, $\psi^2_{m,n}$ is asymptotically equivalent, when H_n is true, to

$$\sum_{\lambda=1}^{m-1} \lambda M_{m+1-\lambda,m-1-\lambda} + mL_{n+1,n},$$

while $L_{m,n}$ is asymptotically equivalent to $\sum_{\lambda=1}^{m-1} M_{m+1-\lambda,m-1-\lambda} + L_{n+1,n}$. For $n > 0$, the relation between $\psi^2_{m,n}$ and the likelihood ratio statistics $L_{m,n}$ and $M_{m,n}$ is not so straightforward. However, a modification $\psi'^2_{m,n}$ of $\psi^2_{m,n}$ (see Section 6 herein) is asymptotically equivalent, when H_n is true, to $L_{m,n}$ for $m = n + 1$, and to $\sum_{\lambda=1}^{m-n-1} \lambda M_{m+1-\lambda,m-1-\lambda} + (m - n)L_{n+1,n}$ for $m > n + 1$; while the likelihood ratio statistic $L_{m,n}$ is asymptotically equivalent to

$$\sum_{\lambda=1}^{m-n-1} M_{m+1-\lambda,m-1-\lambda} + L_{n+1,n}.$$

In [10], the m-tuple u was "split" into an $(m - n - 1)$-tuple, an n-tuple, and a 1-tuple; thus obtaining s^n "contingency tables" ($n \geqq 0$) each $s^{m-n-1} \times s$ (see [10]). The statistic $M_{m,n}$ can be seen to be asymptotically equivalent to the sum of the "likelihood ratio statistics" (for testing "independence" in each table) for the s^n tables, and the asymptotic distribution, when H'_n is true, of $M_{m,n}$ will be χ^2 with $s^n(s^{m-n-1} - 1)(s - 1) = s^m - s^{m-1} - s^{n+1} + s^n$ degrees of freedom. It is also possible to "split" the m-tuple u into an $(m - n - 1 - r)$-tuple, an n-tuple, and a $(1 + r)$-tuple $(0 \leqq r \leqq m - n - 2)$; thus obtaining s^n "contingency tables," each $s^{m-n-1-r} \times s^{1+r}$ (see [10]). The sum ${}_rM_{m,n}$ of the likelihood ratio (or any equivalent goodness of fit) statistics for the s^n tables will have an asymptotic mean value, when H'_n is true, of

$$s^n(s^{m-n-1-r} - 1)(s^{1+r} - 1) = s^m - s^{m-r-1} - s^{n+1+r} + s^n.$$

but the asymptotic distribution will not be χ^2 unless $r = 0$ or $m - n - 2$. It can be seen, using the methods developed in the present paper, that the statistic ${}_rM_{m,n}$ will be asymptotically equivalent, when H'_n is true, to

$$\sum_{\lambda=1}^{m-n-1} h(\lambda) M_{m+1-\lambda, m-1-\lambda},$$

where

$$h(\lambda) = \begin{cases} \lambda & \text{for } 0 < \lambda \leqq v \\ v & \text{for } v \leqq \lambda \leqq m - n - v \\ (m - n - \lambda) & \text{for } m - n - v \leqq \lambda \leqq m - n - 1, \end{cases}$$

and $v = \min\ [r + 1,\ m - n - r - 1]$. Thus, the asymptotic distribution $(N \to \infty)$ of ${}_rM_{m,n}$ (or the corresponding asymptotically equivalent goodness of fit statistics), when H'_n is true, is

$$\mathop{*}_{\lambda=1}^{m-n-1} K_{g(\lambda)}[x/(h(\lambda))].$$

This result generalizes the earlier published results concerning the asymptotic distribution of the likelihood ratio statistic $M_{m,n}$ (or the corresponding asymptotically equivalent goodness of fit statistics) for testing the null hypothesis H'_n within H'_{m-1}, since ${}_rM_{m,n}$ for $r = 0$ or $m - n - 2$ is asymptotically equivalent to $M_{m,n}$ (see [6], [10]). A proof of this result will not be given since the method of proof is quite similar to that presented here for the asymptotic distribution of $\psi^{*2}_{m,n}$.

2. The Case $n = -1$. Let us first consider the case of equiprobable or perfect randomness $(n = -1)$. We have that $H'_{-1} = H_{-1} = H^*_{-1}$, and $\psi^2_{m,-1} = \psi^{*2}_{m,-1}$. Thus, Conjectures (a) and (b) must be in agreement when $n = -1$. For $n = -1$, Conjecture (a) states that the asymptotic distribution of $\psi^{*2}_{m,-1}$ is

$$\mathop{*}_{\lambda=1}^{m} K_{g(\lambda)}(x/\lambda),$$

while (b) states that the asymptotic distribution of $\psi^2_{m,-1}$ is

$$\mathop{*}_{\lambda=1}^{m-1} K_{g(\lambda)}(x/\lambda) * K_{s-1}(x/m).$$

Thus, we must define $K_{g(m)}(x/m)$ as $K_{s-1}(x/m)$; i.e., $K_{(s-1)s^{m-1}}(x/m)$ as $K_{s-1}(x/m)$. It should also be mentioned that $\psi^2_{m,n}$ and $\psi^{*2}_{m,n}$ are defined only for $m \geqq n + 1$ (with $m \geqq 1$, for $n = -1$), and the symbol $\mathop{*}_{\lambda=1}^{0} K$ is to be understood as the atomic distribution that has the total probability 1 at the value $x = 0$. Since H'_{-1} is a special case of H'_0, results for $n = -1$ will follow directly from results for $n = 0$.

3. The Case $n = 0$. In the present paper, it will be convenient deal to with circular sequences, so that (for $m = 2$) $\sum_j f_{ij} = \sum_j f_{ji} = f_i$. A more general statement (for $m \geqq 2$) can be seen to hold for circular sequences (see [6]). A method of modifying results obtained for circular sequences so that they can be applied to linear sequences has been given in [9], and this method can be used to indicate that results analogous to those presented in the present paper will hold for linear sequences. The reader is cautioned that formulas for circular sequences can not be applied directly to linear sequences (see [9] and Corrigenda to [6]). It will also be convenient herein to replace $\psi^2_{m,n}$ and $\psi^{*2}_{m,n}$ by their asymptotically equivalent forms (when H_n is true)

$$\psi^2_{m,n} \approx 2 \sum_u f_u \log (f_u / f_{u,n}),$$

and $\psi^{*2}_{m,n} \approx 2 \sum_u f_u \log (f_u / f^*_{u,n})$, respectively.

Let us first consider Conjecture (a) when $n = 0$. For $m = 1$, this conjecture is obviously correct. For $m = 2$, this conjecture was first stated in [8] and was proved by Dawson and Good [4] and by Goodman [10]. The analogous result for the asymptotically equivalent form of $\psi^{*2}_{2.0}$ was proved by Hoel [11].

For $m = 3$, Conjecture (a) states that

$$\psi^{*2}_{3.0} = \sum_{ijk} (f_{ijk} - f_i f_j f_k/N^2)^2/(f_i f_j f_k/N^2)$$

$$\approx \chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2},$$

where the symbols χ^2_i denote independent random variables each having a chi-square distribution with i degrees of freedom. (The $f_i f_j f_k/N^2$ used above is not the exact expected value, but is an asymptotic approximation; such asymptotic approximations for expected values will be used throughout.) We have that

$$\psi^{*2}_{3.0} \approx 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_i f_j f_k/N^2)]$$

$$= 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_i f_{jk}/N)] + 2 \sum_{jk} f_{jk} \log [f_{jk}/(f_j f_k/N)].$$

The second term in the sum is asymptotically $\chi^2_{(s-1)^2}$, by the result for $m = 2$. The first term in the sum can be split into two parts, thus obtaining

$$2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_{ij} f_{jk}/f_j)] + 2 \sum_{ij} f_{ij} \log [f_{ij} /(f_i f_j/N)].$$

By the results in [10], for the test of H'_1 within H'_2, the asymptotic distribution of the first part is $\chi^2_{s(s-1)^2}$; the asymptotic distribution of the second part is $\chi^2_{(s-1)^2}$ (by the results for $m = 2$). The first part is asymptotically independent of the second part. This can be seen from the fact that their sum has the same asymptotic behavior, under H'_0, as the standard likelihood ratio statistic used in testing independence in an $s^2 \times s$ contingency table (see the test of H'_0 within H'_2 in [10]), and the two parts in the sum are obtained in the same manner as the partitioning of the likelihood ratio for the contingency table into two independent parts (see p. 439 in [3] and the articles referred to therein; rigorous

proofs of some of the published results concerning partitioning of contingency tables are given in [12][2]). The first part is obtained by separating the s^2 rows into s sets of s rows, thus obtaining s contingency tables, each $s \times s$, and using the combined likelihood ratio for the s tables to obtain asymptotically a $\chi^2_{s(s-1)^2}$ distribution (which leads to a test of H_1' within H_2' in [10]); the second part is obtained by combining the s rows in each set to obtain an $s \times s$ contingency table, and using the likelihood ratio for this table to obtain asymptotically a $\chi^2_{(s-1)^2}$ distribution (which leads to a test of H_0' within H_1' in [10]). Since the second part of the first term in $\psi^{*2}_{3.0}$ is equal to the second term in $\psi^{*2}_{3.0}$, their sum is asymptotically $2\chi^2_{(s-1)^2}$. Thus we have proved that $\psi^{*2}_{3.0} \approx \chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2}$

For $m = 4$, Conjecture (a) states that

$$\psi^{*2}_{4.0} = \sum_{ijkl} (f_{ijkl} - f_i f_j f_k f_l/N^3)^2/(f_i f_j f_k f_l/N^3)$$

$$\approx \chi^2_{s^2(s-1)^2} + 2\chi^2_{s(s-1)^2} + 3\chi^2_{(s-1)^2}.$$

We have that

$$\psi^{*2}_{4.0} \approx 2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_i f_j f_k f_l/N^3)]$$

$$= 2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_i f_{jkl}/N)] + 2 \sum_{jkl} f_{jkl} \log [f_{jkl}/(f_j f_k f_l/N^2)]$$

The second term in the sum is $\psi^{*2}_{3.0}$ and is asymptotically $\chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2}$, by Conjecture (a) for $m = 3$. The first term can be split into two parts, thus obtaining

$$2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_{ijk} f_{jkl}/f_{jk})] + 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_i f_{jk}/N)]$$

By the results in [10] for the test of H_2' within H_3', the first part is asymptotically $\chi^2_{s^2(s-1)^2}$; the second part is asymptotically $\chi^2_{(s^2-1)(s-1)}$ (by the results for $m = 3$) The two parts are asymptotically independent. This follows from the fact that their sum has the same asymptotic behavior, under H_0', as the standard likelihood ratio statistic used in testing independence in an $s^3 \times s$ contingency table (see the test of H_0' within H_3' in [10]), and the two parts in the sum are obtained in the same manner as the partitioning of the likelihood ratio for the contingency table into two independent parts. The first part is obtained by separating the s^3 rows into s^2 sets of s rows, thus obtaining s^2 contingency tables, each $s \times s$, and using the combined likelihood ratio for the s^2 tables to obtain $\chi^2_{s^2(s-1)^2}$ (which leads to a test of H_2' within H_3' in [10]); the second part is obtained by combining the s rows in each set to obtain an $s^2 \times s$ contingency table, and using the likelihood ratio for this table we get $\chi^2_{(s^2-1)(s-1)}$ (which leads to a test of H_0' within H_2' in [10]). Since the second part of the first term in $\psi^{*2}_{4.0}$ can be written as $\chi^2_{(s^2-1)(s-1)} = \chi^2_{s(s-1)^2} + \chi^2_{(s-1)^2}$ (see the results for $m = 3$), and since the second term in $\psi^{*2}_{4.0}$ is $\psi^{*2}_{3.0} \approx \chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2}$ (where the $\chi^2_{s(s-1)^2}$ and the $\chi^2_{(s-1)^2}$

[2] *I am indebted to T. W. Anderson for bringing [12] to my attention.*

expressions are identical with those appearing in the second part of the first term), their sum is asymptotically $2\chi^2_{s(s-1)^2} + 3\chi^2_{(s-1)^2}$. We have thus proved that

$$\psi^{*2}_{1.0} \approx \chi^2_{s^2(s-1)^2} + 2\chi^2_{s(s-1)^2} + 3\chi^2_{(s-1)^2}.$$

For $m = 5, 6, \cdots$, the same method of proof applies for Conjecture (a) when $n = 0$; it is easy to see that $\psi^{*2}_{m,0}$ is asymptotically equivalent, under H_0', to a weighted sum of asymptotically independent likelihood ratio statistics.

Let us now consider Conjecture (b) when $n = 0$. We have

$$\begin{aligned}\psi^2_{m,n} &\approx 2 \sum_u f_u \log (f_u/f_{u,n}) \\ &= 2 \{\sum_u f_u \log (f_u/f^*_{u,n}) + \sum_u f_u \log (f^*_{u,n}/f_{u,n})\} \\ &\approx \psi^{*2}_{m,n} + 2 \sum_u f_u \log (f^*_{u,n}/f_{u,n}).\end{aligned}$$

For $m = 1$, $f^*_{u,0} = f_u$, and the second term is $2\sum_i f_i \log (f_i/Np_i)$, which is asymptotically χ^2_{s-1} by the standard statistical theory for goodness of fit tests. For $m = 2$, the second term is

$$\begin{aligned}2 \sum_u f_u \log (f^*_{u,0}/f_{u,0}) &= 2 \sum_{ij} f_{ij} \log [(f_i f_j/N)/Np_i p_j] \\ &= 4 \sum_i f_i \log (f_i/Np_i),\end{aligned}$$

which is asymptotically $2\chi^2_{(s-1)}$. The first term $\psi^{*2}_{2,0}$ is asymptotically independent of the second. This follows from the fact that the sum of $\psi^{*2}_{2.0}$ and $2\sum f_i \log (f_i/Np_i)$ is the likelihood ratio obtained in testing the null hypothesis H_0 that the transition probabilities for the Markov chain are $p_{ij} = p_j = p_j^0$ (specified) within the hypothesis H_1' (i.e., $2\sum_{ij} f_{ij} \log (f_{ij} / f_i p_j) \approx \chi^2_{(s-1)^2} + \chi^2_{(s-1)} = \chi^2_{s(s-1)}$ (see [1])), and the two terms in the sum are obtained by partitioning the likelihood ratio into two independent parts (the independence of the two parts follows directly from an examination of the asymptotic behavior of the f_{ij} (see, e.g., [9])). The first part is asymptotically $\chi^2_{(s-1)^2}$ and tests the null hypothesis H_0' that $p_{ij} = p_j$ (unspecified) within H_1'; the second part is asymptotically $\chi^2_{(s-1)}$ and tests the null hypothesis H_0 that $p_j = p_j^0$ (specified) within H_0'. Thus, $\psi^2_{2,0} \approx \chi^2_{(s-1)^2} + 2\chi^2_{(s-1)}$.

For $m = 3$ the second term is

$$\begin{aligned}2 \sum_u f_u \log(f^*_{u,0}/f_{u,0}) &= 2 \sum_{ijk} f_{ijk} \log [(f_i f_j f_k/N^2)/Np_i p_j p_k] \\ &= 6 \sum_i f_i \log (f_i/Np_i),\end{aligned}$$

which is asymptotically $3\chi^2_{(s-1)}$. The first term is independent of the second, by a similar argument to that presented for $m = 2$. Thus,

$$\psi^2_{3,0} \approx \chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2} + 3\chi^2_{(s-1)}.$$

For $m = 4, 5, 6, \cdots$, the same method of proof applies for Conjecture (b) when $n = 0$.

We have thus given an altogether different method for proving the results obtained in [2] for $n = 0$; the results in [2] were based on the theory of finite-dimensional vector spaces. Since H'_{-1} is a special case of H'_0, the results given in the present section also prove that Conjectures (a) and (b), when properly interpreted, are true for $n = -1$, which generalizes the result proved in [5] for $n = -1$ and s prime. The different method presented in the present paper may further the understanding of the results in [2] and [5].

4. The Case $n = 1$. Let us now consider Conjecture (a) when $n = 1$. For $m = 2$, the conjecture is obviously true. For $m = 3$, we have that

$$\psi_{3.1}^{*2} \approx 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_{ij} f_{jk}/f_j)] \approx \chi^2_{s(s-1)^2},$$

by the results in [10] for the test of H'_1 within H'_2 For $m = 4$, we have that

$$\psi_{4\,1}^{*2} \approx 2 \sum_{ijkl} f_{ijkl} \log \{f_{ijkl}/[f_{ij} f_{jk} f_{kl}/(f_j f_k)]\}$$

$$= 2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_{ijk} f_{jkl}/f_{jk})] + 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_{ij} f_{jk}/f_j)]$$

$$+ 2 \sum_{jkl} f_{jkl} \log [f_{jkl}/(f_{jk} f_{kl}/f_k)].$$

By the results in [10] for the test of H'_2 within H'_3, the first term in the sum is asymptotically $\chi^2_{s^2(s-1)^2}$, and the second term is asymptotically $\chi^2_{s(s-1)^2}$ (see $m = 3$). The first term is asymptotically independent of the second This follows from the fact that their sum can be regarded as the combined likelihood ratio used in testing independence in s contingency tables, each $s^2 \times s$ (see the test of H'_1 within H'_3 in [10]), and the two terms in the sum are obtained by partitioning the likelihood ratio for each of the s tables into two independent parts. For each of the s tables, the first part is obtained by separating the s^2 rows into s sets of s rows, thus obtaining s new tables, each $s \times s$, and using the combined likelihood ratio for the total of s^2 tables to obtain $\chi^2_{s^2(s-1)^2}$ (which is a test of H'_2 within H'_3 in [10]); the second part, for each of the original s tables, is obtained by combining the s rows in each new table to obtain an $s \times s$ table, and using the likelihood ratio for this table (there are s such tables) we get $\chi^2_{s(s-1)^2}$ (which is a test of H'_1 within H'_2 in [10]). The third term in the sum is asymptotically $\chi^2_{s(s-1)^2}$ (see $m = 3$), and it is equal to the second term in the sum. Thus we have $\psi_{4\,1}^{*2} \approx \chi^2_{s^2(s-1)^2} + 2\chi^2_{s(s-1)^2}$.

For $m = 5, 6, \cdots$, the same method of proof applies for Conjecture (a) when $n = 1$; $\psi_{m\,1}^{*2}$ is asymptotically equivalent to a weighted sum of asymptotically independent likelihood ratio statistics, under H'_1.

Let us now consider Conjecture (b) when $n = 1$. We have that

$$\psi_{m.1}^{2} \approx \psi_{m.1}^{*2} + 2 \sum_{\bullet} f_{\bullet} \log (f^{\bullet}_{\bullet.1}/f_{\bullet.1}).$$

For $m = 2$, $f^*_{u,1} = f_u$; thus, the first term $\psi^{*2}_{m,1} = 0$, and the second term is $2\sum_{ij} f_{ij} \log (f_{ij} / Np_i p_{ij})$, where the p_i are the stationary probabilities for the first order Markov chain with constant transition matrix $P = [p_{ij}]$. Conjecture (b) states that $\psi^2_{2,1} \approx \chi^2_{(s-1)^2} + 2\chi^2_{(s-1)}$. We could write

$$\psi^2_{2,1} \approx 2 \sum_{ij} f_{ij} \log [f_{ij}/(f_i f_j/N)] + 2 \sum_{ij} f_{ij} \log [(f_i f_j/N)/(Np_i p_{ij})].$$

The first is not asymptotically $\chi^2_{(s-1)^2}$, except when $n = 0$; and the second term is not asymptotically $2\chi^2_{(s-1)}$, except when $n = 0$. It is easy to see that Conjecture (b) will not hold true for $n = 1$, nor for $n > 1$.

Conjecture (b) will now be modified and this modified version will be proved true. This modification, for the special case $n = 1$, was first mentioned to the author by P. Billingsley in a private communication. In this communication, he mentioned that he had also obtained independently a proof of Conjecture (a), for the case $n = 1$, by very different methods from those used in the present paper, and that his results for Conjecture (a) and the modified Conjecture (b), when $n = 1$, could be extended to the case when $n > 1$, although the detailed asymptotic distributions were not given in the more general case [13].

Let $\psi'^2_{m,1} = \sum_u (f_u - f'_{u,1})^2/f'_{u,1}$, where $f'_{u,1}$ is the expected value of f_u in a new sequence of length N given H_1 and f_{u_1}; i.e., $f'_{u,1} = f_{u_1}\Pi_{i=1}^{m-1} p_{u_i u_{i+1}}$. Then

$$\psi'^2_{m,1} \approx \psi^{*2}_{m,1} + 2 \sum_u f_u \log (f^*_{u,1}/f'_{u,1}).$$

When $m = 2$, the first term $\psi^{*2}_{2,1}$ in the sum is zero and the second term is

$$2 \sum_{ij} f_{ij} \log (f_{ij}/f_i p_{ij}),$$

which is asymptotically $\chi^2_{s(s-1)}$ (see [1]). Thus, the asymptotic distribution of $\psi'^2_{2,1}$ is $\chi^2_{s(s-1)}$.

When $m = 3$, the first term $\psi^*_{3,1}$ is asymptotically $\chi^2_{s(s-1)^2}$, and the second term is

$$2 \sum_{ijk} f_{ijk} \log (f_{ij} f_{jk}/f_j f_i p_{ij} p_{jk}) = 4 \sum_{ij} f_{ij} \log (f_{ij}/f_i p_{ij}),$$

which is asymptotically $2\chi^2_{s(s-1)}$. The first term leads to a test of H'_1 within H'_2, and the second term leads to a test of H_1 within H'_1; it can be seen that the two terms are asymptotically independent under H_1. Thus, for $m = 3$, the asymptotic distribution of $\psi'^2_{m,1}$, when H_1 is true, is

$$\mathop{*}_{\lambda=1}^{m-2} K_{g(\lambda)}(x/\lambda) * K_{s(s-1)}[x/(m - 1)].$$

This result can be proved for $m \geqq 3$ by the same method as given here for $m = 3$. Thus, a modified version of Conjecture (b) holds true for $n = 1$.

5. The Case $n = 2$. Let us now consider Conjecture (a) when $n = 2$. For $m = 3$, the conjecture is obviously true. For $m = 4$, we have

$$\psi^{*2}_{4,2} \approx \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_{ijk} f_{jkl}/f_{jk})] \approx \chi^2_{s^2(s-1)^2},$$

by the results in [10]. For $m = 5$, we have

$$\psi_{5,2}^{*2} \approx 2 \sum_{ijklm} f_{ijklm} \log [f_{ijklm}/(f_{ijk} f_{jkl} f_{klm}/f_{jk} f_{kl})]$$

$$= 2 \sum_{ijklm} f_{ijklm} \log [f_{ijklm}/(f_{ijkl} f_{jklm}/f_{jkl})] + 2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_{ijk} f_{jkl}/f_{jk})]$$

$$+ 2 \sum_{jklm} f_{jklm} \log [f_{jklm}/(f_{jkl} f_{klm}/f_{kl})].$$

By the results in [10], the first term in the sum is asymptotically $\chi^2_{s^3(s-1)^2}$, and the second term is asymptotically $\chi^2_{s^2(s-1)^2}$ (see $m = 4$). The first term is asymptotically independent of the second, this follows by an argument similar to those appearing earlier here. The third term in the sum is asymptotically

$$\chi^2_{s^2(s-1)^2}$$

(see $m = 4$), and it is equal to the second term in the sum. Thus, we have $\psi_{5,2}^{*2} \approx \chi^2_{s^3(s-1)^2} + 2\chi^2_{s^2(s-1)^2}$.

For $m = 6, 7, \cdots$, the same method of proof applies for Conjecture (a) when $n = 2$. Conjecture (b) will not be true for $n = 2$, as it was not for $n = 1$ A modification of Conjecture (b) for $n = 2$ will now be given, which is similar to, although different from, Billingsley's modification of this conjecture for the special case $n = 1$.

Let $\psi_{m,2}'^2 = \sum_{\mathbf{u}} (f_{\mathbf{u}} - f'_{\mathbf{u},2})^2/f'_{\mathbf{u},2}$, where $f'_{\mathbf{u},2}$ is the expected value of $f_{\mathbf{u}}$ in a new sequence of length N given H_2 and $f_{u_1u_2}$, i e, $f'_{\mathbf{u},2} = f_{u_1u_2} \prod_{i=1}^{m-2} p_{u_iu_{i+1}u_{i+2}}$ where $p_{u_1u_2u_3}$ is the second order transition probability that $X_t = u_3$, given that $X_{t-1} = u_2$ and $X_{t-2} = u_1$. Then $\psi_{m,2}'^2 \approx \psi_{m,2}^{*2} + 2 \sum_{\mathbf{u}} f_{\mathbf{u}} \log (f^*_{\mathbf{u},2}/f'_{\mathbf{u},2})$. When $m = 3$, the first term $\psi_{3,2}^{*2}$ in the sum is zero, and the second term is $2 \sum_{ijk} f_{ijk} \log (f_{ijk}/f_{ij}p_{ijk})$, which is asymptotically $\chi^2_{s^2(s-1)}$ (see [1]). Thus, the asymptotic distribution of $\psi_{3,2}'^2$ is $\chi^2_{s^2(s-1)}$.

When $m = 4$, the first term $\psi_{4,2}^{*2}$ is asymptotically $\chi^2_{s^2(s-1)^2}$, and the second term is

$$2 \sum_{ijkl} f_{ijkl} \log (f_{ijk} f_{jkl}/f_{jk} f_{ij} p_{ijk} p_{jkl}) = 4 \sum_{ijk} f_{ijk} \log (f_{ijk}/f_{ij} p_{ijk}),$$

which is asymptotically $2\chi^2_{s^2(s-1)}$. The first term leads to a test of H_2' within H_3', and the second term leads to a test of H_2 within H_2'; it can be seen that the two terms are asymptotically independent under H_2. Thus, for $m = 4$, the asymptotic distribution of $\psi_{m,2}'^2$, when H_2 is true, is

$$\mathop{*}_{\lambda=1}^{m-2} K_{g(\lambda)}(x/\lambda) * K_{s^2(s-1)}[x/(m-2)].$$

This result can be proved for $m \geq 4$ by the same method as given here for $m = 4$ Thus, a modified version of Conjecture (b) holds true for $n = 2$.

6. The General Case. The method of proof used in the preceding sections for $n = -1, 0, 1, 2$ can also be applied when $n = 3, 4, \cdots$. In this way, Conjecture (a) can be proved in the general case $n \geq -1$ and the following modification of

Conjecture (b) also holds in the general case. Let $\psi'^2_{m,n} = \sum_u (f_u - f'_{u,n})^2/f'_{u,n}$, where $f'_{u,n}$ is the expected value of f_u in a new sequence of length N given H_n and $f_{u_1 u_2 \cdots u_n}$ $(n \geqq 1)$. Then, the asymptotic distribution of $\psi'^2_{m,n}$, when H_n is true, is

$$\mathop{*}_{\lambda=1}^{m-n-1} K_{g(\lambda)}(x/\lambda) * K_{s^n(s-1)}[x/(m-n)].$$

If we define $\psi'^2_{m,0}$ as $\psi^2_{m,0}$, then Conjecture (b) for $n = 0$, is identical with the modified version, and it also holds true. For $n = -1$, H_{-1} is a special case of H'_0, and the modified version of Conjecture (b) can be applied with n taken as zero. The reader will note that the asymptotic distribution of $\psi'^2_{m,n}$ is not mathematically independent of n; neither was the asymptotic distribution of $\psi^{*2}_{m,n}$. The result presented here for $\psi'^2_{m,n}$ generalizes Billingsley's result for $n = 1$.

A direct proof of these results could be given for the general case; this was not done here, since the proof proceeds along the same lines as the earlier discussion herein, and the results may be simpler to understand by considering first $n = 0$, $m = 1, 2, 3, 4, \cdots$; $n = 1$, $m = 2, 3, 4, \cdots$; $n = 2$, $m = 3, 4, \cdots$; etc.

In closing, we mention another conjecture by I. J. Good. In [6], the author conjectures that, when H'_{m-1} is true, the variables $-2 \log \lambda_{m-1,m}$ $(m = 0, 1, 2, \cdots)$ are asymptotically independent, where $\lambda_{m-1,m}$ is the ratio of the maximum likelihood given H'_{m-1} to that given H'_m. If this conjecture were true, than an elegant proof of some results for testing H'_m within H'_n would be available (see [6]). We have that $-2 \log \lambda_{m-1,m} \approx \psi^{*2}_{m+1,m-1}$, when H'_{m-1} is true. The asymptotic independence of the likelihood ratios follows by the same kind of argument presented earlier in the present paper for the independence of some of the statistics considered (see, e.g., the reason why $\psi^{*2}_{4,2}$ and $\psi^{*2}_{3,1}$ are asymptotically independent, given $n = 1$, in the discussion here of the case $m = 4$ and $n = 1$).

The reader is referred to [13] for results that are closely related to some of those presented here, although the general approach and methods are very different. Also, some of the work in [14], [15], and [16] has some (but not much) relation to the present paper.

REFERENCES

[1] M. S. Bartlett, "The frequency goodness of fit test for probability chains," *Biometrika*, Vol. 47 (1951), pp. 86–95.

[2] P. Billingsley, "Asymptotic distributions of two goodness of fit criteria," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1123–9.

[3] W. G. Cochran, "Some methods for strengthening the common χ^2 tests," *Biometrics*, Vol. 10 (1954), pp. 417–51.

[4] R. Dawson and I. J. Good, "Exact Markov probabilities from oriented linear graphs," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 838–946.

[5] I. J. Good, "The serial test for sampling numbers and other tests for randomness," *Proc. Cambridge Philos. Soc.*, Vol. 49 (1953), pp. 276–84.

[6] I. J. Good, "The likelihood ratio test for Markoff chains," *Biometrika*, Vol. 42 (1955), pp. 531–3; Corrigenda, Vol. 44 (1957), p. 301.

[7] I. J. Good, "Review of P. Billingsley's 'Asymptotic distributions of two goodness of fit criteria'," *Mathematical Reviews*, Vol. 18 (1957), p. 607.

[8] LEO A. GOODMAN, "On the statistical analysis of Markov chains" (abstract), *Ann. Math. Stat.*, Vol. 26 (1955), p. 771

[9] LEO A. GOODMAN, "Simplified runs tests and likelihood ratio tests for Markoff chains," *Biometrika*, Vol. 45 (1958), pp. 181–97.

[10] LEO A. GOODMAN, "Exact probabilities and asymptotic relationships for some statistics from m-th order Markov chains," *Ann Math Stat.*, Vol. 29 (1958), pp. 476–90.

[11] P. G. HOEL, "A test for Markoff chains," *Biometrika*, Vol 41 (1954), pp. 430–3.

[12] MARILYN SPENCER, "The two-way contingency table and approximate chi-square distributions," Master's Essay, Columbia University Library, 1957.

[13] P. BILLINGSLEY, "On testing Markov Chains," unpublished report.

[14] G P. BASHARIN, "On using the test of goodness of fit (χ^2 agreement) as a test of independence of trials," *Teoriya Veroyatænostei i ee Primereniya*, Vol. 2 (1957), pp. 141–2.

[15] N. V. SMIRNOV, "The statistical estimation of transition probabilities in Markov chains," *Vestnik Leningradskoyo Universitcta*, Vol. 1 (1955), pp 47–8.

[16] V. E. STEPANOW, "Some statistical tests for Markov chains," *Teoriya Veroyatnostei i ee Primeneniya*, Vol. 2 (1957), pp 143–4

EFFICIENCY PROBLEMS IN POLYNOMIAL ESTIMATION

By Paul G. Hoel

University of California, Los Angeles

1. Summary. Using the generalized variance as a criterion for the efficiency of estimation, the best choice of fixed variable values within an interval for estimating the coefficients of a polynomial regression curve of given degree is determined for the classical regression model. Using this same criterion, some results are obtained on the increased efficiency arising from doubling the number of equally spaced observation points

(i) when the total interval is fixed and

(ii) when the total interval is doubled. Measures of the increased efficiency are found for the classical regression model and for models based on a particular stationary stochastic process and a pure birth stochastic process.

2. Introduction. In the classical theory of regression, a set of values $x_1, x_2, \cdots, x_n$ of a variable x is selected and observations are made on a related variable y corresponding to those selected x values. If y_i denotes the y value corresponding to x_i, it is then assumed that $y_1, y_2, \cdots, y_n$ are uncorrelated variables with a common variance σ^2. Now if it is assumed that the means of the y's lie on a polynomial curve of degree k, that is, that

$$E(y_i) = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k \tag{1}$$

then a basic problem in statistics is how best to estimate the β's.

There are two aspects to this estimation problem. One is to determine the best method for using the information given by a set of n observtions $y_1, y_2, \cdots, y_n$. The other is to determine the best method for choosing the x values at which to take observations.

Although much research has gone into studying the first aspect of the problem, considerably less has been done on the second. Many years ago, K. Smith [1] was able to determine those x values within a fixed interval that minimize the maximum variance of a single estimated ordinate for polynomials up to degree six. More recently, De La Garza [2] was able to show that just as much information is obtained from observations made at certain $k + 1$ points in the interior of an interval as from n distinct points in that interval. Elfving [3], Chernoff [4], Daniels [5], and Ehrenfeld [6] have also made contributions toward this and other closely related problems.

In this paper an optimum solution based on the generalized variance is given for the problem of how to choose the x values in an interval for the classical regression model. In addition, a beginning is made on the more general problem

Received January 24, 1958; revised June 18, 1958.

of how to choose x values for efficient polynomial estimation when one drops the assumption that the y's are uncorrelated.

3. Estimation methods. When a number of parameters are to be estimated simultaneously, the volume of the ellipsoid of concentration of the estimates is often used as a measure of the efficiency of the estimates. Since the square of the volume of the ellipsoid of concentration is proportional to the generalized variance of the estimates, one can just as well use the generalized variance as a measure of efficiency. This is the measure that will be used in this paper for making comparisons of different sets of estimates.

Suppose one wished to estimate the function $\lambda_0\beta_0 + \lambda_1\beta_1 + \cdots + \lambda_k\beta_k$, where the λ's are an arbitrary set of real numbers, by means of a linear estimate, $c_1y_1 + c_2y_2 + \cdots + c_ny_n$. Suppose further that the estimate is to be unbiased and possess minimum variance. Then it can be shown that the resulting estimates for the β's are given by the matrix formula

$$\hat{\beta} = (X'S^{-1}X)^{-1}X'S^{-1}y \tag{2}$$

where S is the covariance matrix of the y's and X is the matrix

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdot & x_1^k \\ 1 & x_2 & x_2^2 & \cdot & x_2^k \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdot & x_n^k \end{bmatrix}$$

Furthermore, it can also be shown that the generalized variance of these estimates is given by the determinant formula

$$\text{G.V.} = |X'S^{-1}X|^{-1} \tag{3}$$

These same formulas will be obtained if one assumes that the y's possess a multivariate normal distribution and then finds the maximum likelihood estimates of the β's.

The advantage of the estimates given by formula (2) lies in the fact that it can be shown that among all linear unbiased estimates of the β's, the estimates given by this formula possess a minimum generalized variance. Thus, if one restricts himself to linear estimates, these are optimum estimates. All the comparisons to be made in the following sections will assume that the estimates are those given by formula (2), and hence that the generalized variance is given by formula (3).

4. Classical regression. Since the classical regression model assumes that $y_1, y_2, \cdots, y_n$ are uncorrelated with a common variance σ^2, the covariance matrix S is a diagonal matrix with elements σ^2.

Now De La Garza [2] has shown that the same information matrix, $X'S^{-1}X$, and hence the same value of the generalized variance, can be obtained by replacing a given set of n observations at the points $x_1, x_2, \cdots, x_n$ by a total of n

observations made at $k + 1$ properly selected points in the interval from x_1 to x_n. These points will be denoted by $t_1, t_2, \cdots, t_{k+1}$ and the number of observations to be made at t_i will be denoted by n_i, where $\sum_{i=1}^{k+1} n_i = n$. In terms of these substitute observations, the matrices in (3) are all square matrices and therefore the determinant of their product can be obtained by taking the product of their determinants. As a result, (3) will assume the form

$$\frac{1}{\text{G.V.}} = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_{k+1} \\ \vdots & \vdots & & \vdots \\ t_1^k & t_2^k & \cdots & t_{k+1}^k \end{vmatrix} \begin{vmatrix} \frac{n_1}{\sigma^2} & 0 & \cdots & 0 \\ 0 & \frac{n_2}{\sigma^2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{n_{k+1}}{\sigma^2} \end{vmatrix} \begin{vmatrix} 1 & t_1 & \cdots & t_1^k \\ 1 & t_2 & \cdots & t_2^k \\ \vdots & \vdots & & \vdots \\ 1 & t_{k+1} & \cdots & t_{k+1}^k \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_{k+1} \\ \vdots & \vdots & & \vdots \\ t_1^k & t_2^k & \cdots & t_{k+1}^k \end{vmatrix}^2 \frac{\prod_{i=1}^{k+1} n_i}{\sigma^{2k+2}}.$$

But this determinant is a Vandermonde determinant with value $\prod_{i<j} (t_i - t_j)$; consequently

$$\text{(4)} \qquad \frac{1}{\text{G.V.}} = \frac{1}{\sigma^{2k+2}} \prod_{i<j}^{k+1} (t_i - t_j)^2 \prod_{i=1}^{k+1} n_i$$

Since $\prod_{i=1}^{k+1} n_i$, subject to the restriction $\sum_{i=1}^{k+1} n_i = n$, is maximized when $n_1 = n_2 = \cdots = n_{k+1}$, it follows that the generalized variance will be minimized for a fixed set of values when the same number of observations is taken at each of the t values. This assumes that n will be chosen to make $n/(k + 1)$ an integer.

Now consider the maximization of $\prod_{i<j} (t_j - t_j)^2$, subject to the restriction that $x_1 \leqq t_i \leqq x_n$, $i = 1, \cdots, k + 1$. If x is transformed linearly so that this restriction assumes the form $-1 \leqq t_i \leqq 1$, $i = 1, \cdots, k + 1$, then it is known [7] that the set of t values that maximizes $\prod_{i<j} (t_i - t_j)^2$ is given by the zeros of a polynomial which is the integral of one of the Legendre polynomials. These zeros can be obtained from the proper tables [8].

It is clear from inspecting the function $\prod (t_i - t_j)^2$ that the end points of the interval will always be chosen as two of the t values. It is also clear that the greater the range of x values, the smaller will be the generalized variance.

In view of the preceding results, it follows that optimum linear estimates of the coefficients of classical polynomial regression are obtained by using the estimates given by formula (2), choosing as large a range of x values as possible, taking observations at the $k + 1$ points in this range given by means of the zeros of a tabulated polynomial, and repeating the experiment as many times as the total set, n, of observations will permit, with n chosen to make $n/(k + 1)$ an integer.

The preceding optimum manner of choosing x values assumes that the generalized variance of the estimates of the coefficients of the regression polynomial is the proper measure of efficiency to use. If the sample regression polynomial curve is to be used exclusively for estimating ordinates of the theoretical regression polynomial curve, then one might prefer a measure of efficiency based on the variances and covariances of such estimated values. From this point of view, let $\tau_1, \cdots, \tau_{k+1}$ denote $k + 1$ arbitrary points chosen in the given interval. Further, let α_i and $\hat{\alpha}_i$ denote the ordinate, and its estimate, of the polynomial regression curve at τ_i. Thus,

$$\alpha_i = \beta_0 + \beta_1\tau_i + \cdots + \beta_k\tau_i^k, \qquad i = 1, \cdots, k + 1$$

and

$$\hat{\alpha}_i = \hat{\beta}_0 + \hat{\beta}_1\tau_i + \cdots + \hat{\beta}_k\tau_i^k, \qquad i = 1, \cdots, k + 1$$

Calculations will yield the covariance formula

$$m_{ij} = E(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j) = \sum_{r=0}^{k}\sum_{s=0}^{k} \sigma_{rs}\tau_i^r\tau_j^s$$

where σ_{rs} is the covariance of $\hat{\beta}_r$ and $\hat{\beta}_s$. Since the generalized variance is the determinant of the covariance matrix, the generalized variance of the $\hat{\alpha}$'s will be equal to the determinant $|m_{ij}|$. But it will be observed that the matrix (m_{ij}) can be written in the form

$$(m_{ij}) = \begin{bmatrix} 1 & \tau_1 & \cdots & \tau_1^k \\ 1 & \tau_2 & \cdots & \tau_2^k \\ \vdots & \vdots & & \\ 1 & \tau_{k+1} & \cdots & \tau_{k+1}^k \end{bmatrix} \begin{bmatrix} \sigma_{00} & \cdot & \sigma_{0k} \\ \vdots & & \vdots \\ \sigma_{k0} & \cdot & \sigma_{kk} \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \tau_1 & \tau_2 & \cdots & \tau_{k+1} \\ \vdots & \vdots & & \vdots \\ \tau_1^k & \tau_2^k & \cdots & \tau_{k+1}^k \end{bmatrix}$$

Since $|\sigma_{rs}|$ is the generalized variance of the β's, it follows that

$$\text{G.V. } (\hat{\alpha}) = \text{G.V. } (\hat{\beta}) \prod_{i<j}^{k+1} (\tau_i - \tau_j)^2$$

This result shows that the generalized variance of the estimates of the ordinates of a polynomial regression curve at $k + 1$ arbitrary points in an interval will be minimized when the generalized variance of the estimates of the coefficients of the polynomial regression curve is minimized[1].

A recent paper by Guest [11], which was published after this paper had been submitted, has generalized the results of Smith [1] to polynomials of any degree. He shows that the values of $t_1, t_2, \cdots, t_n$ that minimize the maximum variance of a single estimated ordinate are given by means of the zeros of the derivative of a Legendre polynomial. It is easily seen that this set of values is the same set which minimizes the generalized variance above. Thus, whether one is interested

[1] I am indebted to Professor John Tukey for suggesting this relationship

in efficient estimation of regression coefficients, or in efficient ordinate estimation, either at $k + 1$ points or one point, the optimum choice of t values is the same.

5. Comparison methods. When the assumption that the y's are uncorrelated is dropped, the problem of how best to choose the x's becomes very difficult. The choice will depend in a complicated manner upon the covariance matrix S. As a consequence, comparisons will be made only for equally spaced sets of points and only for three classes of covariance matrices. The sets of points that were selected for consideration are the following:

(1) n equally spaced points in the interval $(0, l)$
(2) $2n$ equally spaced points in the interval $(0, l)$
(3) $2n$ equally spaced points in the interval $(0, 2l)$
(4) two sets of observations of type (1).

A comparison of the relative advantages of choices (2), (3), and (4) over (1) will be made by comparing their generalized variances. Letting δ denote the interval between consecutive x values, these generalized variances will be denoted by G.V. (n, δ), G.V. $(2n, \delta/2)$, G.V. $(2n, \delta)$, and G.V. (2 runs), respectively.

The three classes of covariance matrices that will be studied are the following:

(a) uncorrelated variables, common variance
(b) $\rho(y_i, y_j) = e^{-\alpha|x_i - x_j|}$, $a > 0$, common variance
(c) covariance matrix of a pure birth stochastic process.

The first of these is the classical regression model considered in the preceding section. The second is the covariance matrix of a particular stationary stochastic process. The third was selected because it represents a stochastic process of the non-stationary type and in which the covariances grow as x increases. These three covariance matrices cover a rather wide range of correlation relationships and therefore conclusions obtained from them should have a rather wide range of application.

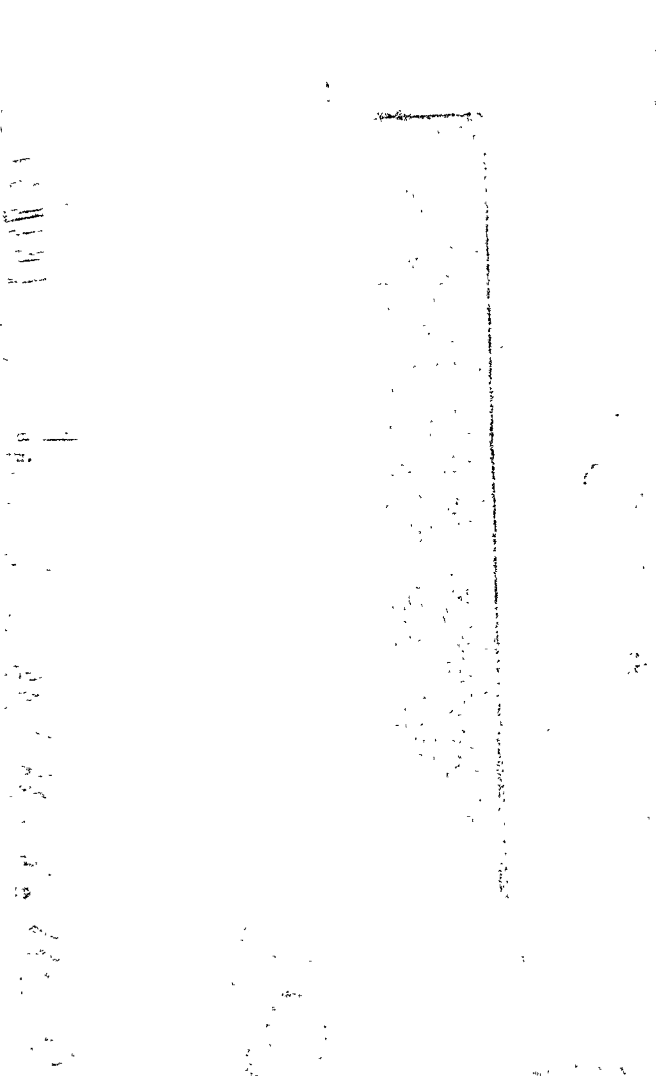
For comparison purposes it is advantageous to consider the following three ratios:

$$R_1 = \left[\frac{\text{G.V. } (n, \delta)}{\text{G.V. } (2n, \delta/2)}\right]^{1/(k+1)}$$

$$R_2 = \left[\frac{\text{G.V. } (n, \delta)}{\text{G.V. } (2n, \delta)}\right]^{1/(k+1)} \tag{5}$$

$$R_3 = \left[\frac{\text{G.V. } (n, \delta)}{\text{G.V. } (m) \text{ runs}}\right]^{1/(k+1)}$$

The reason for these choices is that it is easily shown that R_3 has the value m; consequently if the value of R_1, for example, should turn out to be m, it can be concluded that m runs of the basic experiment are needed to yield the same efficiency of estimation as that obtained by doubling the number of equally spaced observation points in the given interval. All comparisons will be made



in this manner, that is, by stating the number of runs of the experiment needed to yield the same efficiency as the choice of x values being considered.

6. Uncorrelated variables. It will be assumed that $n > k + 1$; consequently the X matrix in (3) will not be a square matrix and formula (4) will not be applicable. Under equal spacing in the interval $(0, l)$, the x values will be chosen as $x_i = i\delta$. As a result, the X matrix will assume the form

$$X = \begin{bmatrix} 1 & \delta & \cdots & \delta^k \\ 1 & 2\delta & \cdots & (2\delta)^k \\ \vdots & \vdots & & \vdots \\ 1 & n\delta & \cdot & (n\delta)^k \end{bmatrix} \tag{6}$$

Since S^{-1} is a diagonal matrix with elements $1/\sigma^2$, it is easily seen that (3) reduces to

$$\frac{1}{\text{G.V.}\ (n, \delta)} = \frac{\delta^{k(k+1)}}{\sigma^{2k+2}} \begin{vmatrix} n & \sum_1^n i & & \sum_1^n i^k \\ \sum_1^n i & \sum_1^n i^2 & & \sum_1^n i^{k+1} \\ \vdots & & & \vdots \\ \sum_1^n i^k & \sum_1^n i^{k+1} & \cdot & \sum_1^n i^{2k} \end{vmatrix}. \tag{7}$$

The value of this determinant is known [10] to be the polynomial displayed in (8); hence

$$\frac{1}{\text{G.V.}\ (n, \delta)} = \frac{\delta^{k(k+1)}}{\sigma^{2k+2}} A n^{k+1}(n^2 - 1^2)^k(n^2 - 2^2)^{k-1} \cdots (n^2 - k^2) \tag{8}$$

where $A = (1!\,2! \cdots k!)^4 / (1!\,2! \cdots (2k + 1)!)$ The value of R_1 given in (5) then becomes

$$R_1 = \frac{1}{2^k}\left[\frac{(2n)^{k+1}(4n^2 - 1^2)^k \cdots (4n^2 - k^2)}{n^{k+1}(n^2 - 1^2)^k \cdots (n^2 - k^2)}\right]^{1/k+1}. \tag{9}$$

Using (8) and (5), it follows readily that

$$R_2 = 2^k R_1. \tag{10}$$

Now consider the limiting values of R_1 and R_2 as $n \to \infty$. The resulting values may be considered as asymptotic measures of efficiency. From (9) and (10) it follows that

$$\lim_{n\to\infty} R_1 = 2 \quad \text{and} \quad \lim_{n\to\infty} R_2 = 2^{k+1}.$$

The first result implies that if one has a large number of equally spaced points in a fixed interval at which observations are made, then two runs of the experi-

ment will yield the same efficiency of estimation as doubling the number of equally spaced points in that interval. The second result implies, for example, that if the polynomial regression curve is of degree 4, then 32 runs of the experiment will be needed to yield the same efficiency of estimation as doubling the number of points by doubling the interval over which observations are to be made. It is clear from this second result that the higher the degree of the polynomial the more important it is to extend the range of x values as far as possible.

7. Stationary process model. Denoting the correlation between y_i and y_j by ρ_{ij}, it follows under equal spacing that the correlation function for model (b) will assume the form

$$\rho_{ij} = e^{-\alpha|x_i - x_j|} = e^{-\alpha\delta|i-j|}$$

Letting $w = e^{-\alpha\delta}$ and setting $\sigma^2 = 1$, since it will always cancel out in the R ratios, it will be seen that the covariance matrix here is given by

$$S = \begin{bmatrix} 1 & w & w^2 & \cdots & w^{n-1} \\ w & 1 & w & \cdots & w^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ w^{n-1} & w^{n-2} & w^{n-3} & \cdots & 1 \end{bmatrix}$$

Calculations will show that the inverse of S is given by

$$S^{-1} = \frac{1}{1-w^2}\begin{bmatrix} 1 & -w & 0 & \cdots & 0 & 0 \\ -w & 1+w^2 & -w & \cdots & 0 & 0 \\ 0 & -w & 1+w^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -w & 1 \end{bmatrix}$$

If S^{-1} is written as the sum of several matrices and then premultiplied by X' and postmultiplied by X, and finally brought together again into one matrix, it will be found that (3) assumes the form

$$\frac{1}{\text{G.V.}(n, \delta)} = \frac{\delta^{k(k+1)}}{(1-w^2)^{k+1}} \mid B(n, w) \mid \tag{11}$$

where $B(n, w)$ is the matrix whose element in row $p + 1$ and column $q + 1$ is given by

$$\begin{aligned} b_{p+1\, q+1} = {} & (w^2 + 1) \sum_1^n i^{p+q} \\ & - w \sum_2^n [i^p(i-1)^q + i^q(i-1)^p] - w^2[n^{p+q} + 1]. \end{aligned} \tag{12}$$

Since $w = e^{-\alpha\delta}$, the value of G.V. $(2n, \delta/2)$ can be obtained by replacing n by $2n$, δ by $\delta/2$, and w by $\sqrt{w}$ in (11). As a result, it will follow that

$$R_1 = \frac{1 + w}{2^k} \frac{| B(2n, \sqrt{w}) |^{1/(k+1)}}{| B(n, w) |^{1/(k+1)}}.$$

Similarly,

$$R_2 = \frac{| B(2n, w) |^{1/(k+1)}}{| B(n, w) |^{1/(k+1)}}.$$

Now allow $n \to \infty$. From (12) it will be observed that the dominating part of $b_{p+1\,q+1}$ is $(w - 1)^2 \sum i^{q+q}$. As a result, the asymptotic value of the determinant $| B(n, w) |$ is

$$\begin{vmatrix} (w-1)^2 n & (w-1)^2 \sum i & (w-1)^2 \sum i^k \\ (w-1)^2 \sum i & (w-1)^2 \sum i^2 & (w-1)^2 \sum i^{k+1} \\ \vdots & \vdots & \vdots \\ (w-1)^2 \sum i^k & (w-1)^2 \sum i^{k+1} \quad \cdot & (w-1)^2 \sum i^{2k} \end{vmatrix}$$

But this is merely $(w - 1)^{2k+2}$ times the determinant in (7), which in turn has the asymptotic value $An^{(k+1)^2}$. From the preceding results, it follows that

$$\lim_{n\to\infty} R_1 = \frac{w + 1}{2^k} \frac{(\sqrt{w} - 1)^2 2^{k+1}}{(w - 1)^2} = \frac{2(w + 1)}{(\sqrt{w} + 1)^2}$$

and

$$\lim_{n\to\infty} R_2 = 2^{k+1}$$

For the purpose of seeing the implications of these formulas, consider the numerical value $w = e^{-\alpha\delta} = .64$. This value implies that the correlation coefficient between neighboring y values is .64. Calculations yield the values

$$\lim_{n\to\infty} R_1 = 1.01 \qquad \text{and} \qquad \lim_{n\to\infty} R_2 = 2^{k+1}.$$

Thus, doubling the number of observation points in a given interval, when there are already a large number of such points, gives practically no additional estimation information. The value of R_2, however, shows that the same asymptotic efficiency is gained here as in the case of uncorrelated variables. For correlated variables like those being considered in this section, it is clear that the interval over which observations are to be made should be extended as far as possible, but that if it can't be extended, repeating the experiment is far more efficient than taking additional observation points.

8. Pure birth process model. Although a pure birth process is a discrete process with an exponential regression curve, it was selected only for its covariance matrix properties which are quite different from those of the two preceding models.

If b denotes the constant asymptotic birth rate, y_0 the population size at time t_0, and y the population size at time $t > t_0$, then the conditional probability function for y, given y_0, is

$$P\{y_0, y; t_0, t\} = \binom{y-1}{y_0-1} e^{-by_0(t-t_0)}[1 - e^{-b(t-t_0)}]^{y-y_0}.$$

Using this formula, expected value calculations will show that the covariance of y_i and y_j, $j \geqq i$, is given by

$$\sigma_{ij} = y_0 e^{b(t_j-t_0)}[e^{b(t_i-t_0)} - 1].$$

Under equal spacing as before, $t_0 = 0$ and $t_i = i\delta$; hence letting $z = e^{b\delta}$,

$$\sigma_{ij} = y_0 z^j(z^i - 1).$$

From this formula it follows that

$$(13) \qquad \sigma_{i,j+m} = z^m \sigma_{ij} \qquad \text{and} \qquad \sigma_{jj} = \frac{z^j(z^j - 1)}{z^i(z^i - 1)} \sigma_{ii}$$

As a result, the covariance matrix S assumes the form

$$S = \begin{bmatrix} \sigma_{11} & z\sigma_{11} & \cdots & z^{n-1}\sigma_{11} \\ z\sigma_{11} & \sigma_{22} & \cdots & z^{n-2}\sigma_{22} \\ \vdots & \vdots & & \vdots \\ z^{n-1}\sigma_{11} & z^{n-2}\sigma_{22} & \cdots & \sigma_{nn} \end{bmatrix}$$

The second of formulas (13) enables this matrix to be expressed as the product of the following two matrices.

$$\begin{bmatrix} \frac{\sigma_{11}}{z-1} & 0 & \cdots & 0 \\ 0 & \frac{\sigma_{22}}{z^2-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_{nn}}{z^n-1} \end{bmatrix} \begin{bmatrix} z-1 & z^2-z & \cdots & z^n - z^{n-1} \\ z-1 & z^2-1 & \cdots & z^n - z^{n-2} \\ \vdots & \vdots & & \vdots \\ z-1 & z^2-1 & \cdots & z^n - 1 \end{bmatrix}$$

Some rather lengthy calculations will show that the inverse matrix is given by

$$S^{-1} = \frac{1}{y_0(z-1)} \begin{bmatrix} z+1 & -z & \cdots & 0 & 0 \\ -1 & z+1 & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & \cdots & z+1 & -z \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{z} & 0 & \cdots & 0 \\ 0 & \frac{1}{z^2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{z^n} \end{bmatrix}$$

Additional lengthy computations, similar to those employed in the preceding section, will show that (3) assumes the form

$$\frac{1}{\text{G.V.}(n, \delta)} \approx \frac{\delta^{k(k+1)} \mid C(n, z) \mid}{y_0^{k+1}(z-1)^k z^n}$$

where $C(n, z)$ is the matrix whose element in row $p + 1$ and column $q + 1$ is given by

$$c_{p+1\,q+1} = \frac{(2^p - 1^p)(2^q - 1^q)}{z} + \frac{(3^p - 2^p)(3^q - 2^q)}{z^2}$$

$$+ \cdots + \frac{(n^p - (n-1)^p)(n^q - (n-1)^q)}{z^{n-1}}$$

For $p = 1$ or $q = 1$, c_{p+1q+1} is defined by $c_{11} = z^n \quad (z - 1) - 1$, and $c_{i1} = c_{1i} = n^{i+1} - 1, i > 1$.

When $n \to \infty$, the elements of this matrix exclusive of those in the first row and first column, converge to functions of z, for $z > 1$ Let

$$g_{pq}(z) = \lim_{n\to\infty} c_{p+1\,q+1}$$

Since, for $z > 1$, c_{11} dominates c_{i1}, $i > 1$, the determinant $|c(n, z)|$ will possess the asymptotic value

$$\frac{z^n}{z-1}\begin{vmatrix} g_{11}(z) & \cdot & g_{1k}(z) \\ \vdots & & \vdots \\ g_{k1}(z) & & g_{kk}(z) \end{vmatrix}$$

For $k < 5$ it has been shown that the preceding determinant has the value

$$\frac{cz^{k(k-1)/2}}{(z-1)^{k^2}}$$

where c depends on k but not on z. Using these results the asymptotic value of *the generalized variance is given by*

$$(11) \qquad \frac{1}{\text{G.V.}(n, \delta)} \approx \frac{\delta^{k(k+1)} cz^{k(k-1)/2}}{y_0^{k+1}(z-1)^{k^2+k+1}}$$

From this result it is easily shown that

$$\lim_{n\to\infty} R_1 = \frac{(\sqrt{z} + 1)^{(k^2+k+1)/(k+1)}}{2^k(\sqrt{z})^{(k^2-k)/(2k+2)}}.$$

Since (11) does not involve n, it follows that

$$\lim_{n\to\infty} R_2 = 1.$$

As a numerical illustration here, let $z = e^{\delta^2} = 10/9$. This value implies that the correlation between y_1 and y_2 is approximately .7 and increases between

neighboring y values as one moves out on the axis. Calculations here yield the following limiting values for R_1.

k	1	2	3	4
$\lim_{n\to\infty} R_1$	1.47	1.32	1.25	1.20

These limiting values of R_1 show that some additional estimation information is gained by doubling the number of points in a fixed interval but that repeating the experiment yields considerably more information. The limiting value of R_2 would seem to indicate that no additional information is gained by extending the interval. This limiting result, however, is not realistic for small samples as will be seen in the next section.

9. Numerical results. Since the asymptotic measures of estimation efficiency obtained in the preceding sections may not be very realistic for small numbers of observations, some numerical computations were made with the assistance of high speed computing equipment. The values of $w = .64$ and $z = 10/9$ used previously were used in these computations. Values of $n = 5$ and $n = 10$ were chosen but only the results for $n = 10$ are given because some of the $n = 5$ values appeared questionable and because there were only moderate differences between the two sets of values. The limiting values of R_1 and R_2 are shown in parentheses adjacent to the computed values. In these computations, adjustments were made in the values of R_1 and R_2 to allow for the fact that doubling the number of points in an interval extends the total interval spanned by the points when the first point is located at $x = \delta$. These adjustments essentially kept the spanned interval unchanged. This was accomplished by replacing $\delta/2$ by $\delta(n - 1)/(2n - 1)$ in the denominator of R_1 and δ by $\delta(2n - 2)/(2n - 1)$ in the denominator of R_2.

k	Model (a)	Model (b)	Model (c)
1	1.90 (2)	1.03 (1.01)	1.43 (1.47)
2	1.81 (2)	1.02 (1.01)	1.25 (1.32)
3	1.72 (2)	1.02 (1.01)	1.19 (1.25)
4	1.64 (2)	1.03 (1.01)	1.18 (1.20)

k	Model (a)	Model (b)	Model (c)
1	3.80 (4)	2.91 (4)	1.83 (1)
2	7.24 (8)	5.01 (8)	1.85 (1)
3	13.76 (16)	8.94 (16)	3.21 (1)
4	26.24 (32)	16.41 (32)	6.15 (1)

It will be observed that the asymptotic values of R_2 are poor approximations for models (b) and (c). These results seem to indicate that in general one should always attempt to extend the range over which observations are to be taken as far as possible and the higher the degree of polynomial the greater is the advantage. They also seem to indicate that if the range can't be extended, it is considerably more efficient to replicate the experiment than double the number of observations, particularly if the variables are strongly correlated.

REFERENCES

[1] K. Smith, "On the standard deviations of adjusted and interpolated values of an observed polynomial function," *Biometrika*, Vol 12 (1918), pp. 1-85

[2] A. De La Garza, "Spacing of information in polynomial regression," *Ann. Math. Stat*, Vol. 25 (1954), pp. 123-30

[3] G. Elfving, "Optimum allocation in linear regression theory," *Ann Math Stat*, Vol. 23 (1952), pp. 255-62.

[4] H. Chernoff, "Locally optimal designs for estimating parameters," *Ann. Math Stat.*, Vol. 24 (1953), pp. 586-602.

[5] H. E. Daniels, "A distribution free test for regression parameters," *Ann. Math Stat.*, Vol. 25 (1954), pp. 499-513.

[6] S. Ehrenfeld, "On the efficiency of experimental designs," *Ann Math. Stat.*, Vol 26 (1955), pp. 247-255.

[7] I. Schur, "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koefficienten," *Math Zeit*, Vol 1 (1918), p. 378.

[8] G Prevost, *Tables et Formules de Fonctions Spheriques*, Gauthier-Villars, Paris, 1933.

[9] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 494.

[10] T. Muir, *Treatise on the Theory of Determinants*, Longmans, Green and Co (1933), p. 440.

[11] P G. Guest, "The spacing of observations in polynomial regression," *Ann Math Stat*, Vol. 29 (1958), pp. 294-299.

ON THE GENERAL CANONICAL CORRELATION DISTRIBUTION

By A. G. Constantine and A. T. James

Division of Mathematical Statistics, C.S.I.R.O.[1]

1. Summary. The paper is divided into two parts:

A. An elementary derivation of Bartlett's results on the distribution of the canonical correlation coefficients using exterior differential forms. Briefly, our method consists of taking the original multivariate normal distribution, transforming to the canonical correlations and other variables, and then integrating out these extraneous variables.

B. A new method of calculating the conditional moments which appear in Bartlett's expansion of this distribution, based on the process of averaging over the orthogonal group. This method allows the calculation of moments of any order.

Part A

2. Introduction. Bartlett [1] obtained the general canonical correlation distribution as a multiple power series in the true canonical correlations ρ_i. In the case of more than one non-zero correlation ρ_i, the coefficients in this expansion depend on the conditional moments of the sample (ordinary) correlations s_i between the pairs of transformed variates representing the true canonical variates, when the sample canonical correlations r_i between the sample canonical variates are fixed.

Bartlett derived his results by a formal generalization of the argument used by Fisher [2] in calculating the distribution of the multiple correlation coefficient. We shall give a new proof of Bartlett's results in a concrete form more suitable for our purposes. Throughout this paper we shall use the concepts of exterior differential forms and alternating products of these forms. The definition and a discussion of these concepts may be found in James [6].

Consider a dependent vector variate with p components and an independent vector variate with $q \geqq p$ components. (Here the terms "dependent" and "independent" are to be understood in the regression sense.) If we take a sample with $n(\geqq p + q)$ degrees of freedom, we may represent it by the $p + q$ column vectors $\xi_1, \xi_2, \cdots, \xi_p$ and $\eta_1, \eta_2, \cdots, \eta_q$, each containing n components. The dependent vector is considered to be a normal variate, and we may distinguish two cases, according as the independent variate is assumed to be (a) a normal variate or (b) a set of fixed vectors in the sample space. In either case we may, without loss of generality, assume the ξ_i and η_j to be the canonical variates (see

Received September 23, 1957; revised March 1, 1958.

[1] Commonwealth Scientific and Industrial Research Organization, University of Adelaide, South Australia.

Hotelling [4]). This means that in case (a) the n components of each vector are standard normal variates with the joint distribution

$$\prod_{i=1}^{p}\left\{(2\pi)^{-n}(1-\rho_i^2)^{-n/2}\exp\left[\frac{-(\xi_i'\xi_i-2\rho_i\xi_i'\eta_i+\eta_i'\eta_i)}{2(1-\rho_i^2)}\right]\prod_{r=1}^{n}d\xi_{ri}\,d\eta_{ri}\right\} \cdot \prod_{j=p+1}^{q}\left\{(2\pi)^{-n/2}\exp[-\eta_j'\eta_j/2]\prod_{r=1}^{n}d\eta_{rj}\right\}. \tag{2.1}$$

In case (b), the non-central means case, we may assume the components of the ξ_i to be independently distributed with unit variance, and the η_j to be vectors lying along the first q co-ordinate axes of the sample space. $\eta_1, \cdots, \eta_p$ may also be identified with the mean vectors of $\xi_1, \cdots, \xi_p$. The joint distribution of the ξ_{ri} is therefore

$$\prod_{i=1}^{p}\left\{(2\pi)^{-n/2}\exp[-(\xi_i'\xi_i-2\xi_i'\eta_i+\eta_i'\eta_i)/2]\prod_{r=1}^{n}d\xi_{ri}\right\}. \tag{2.2}$$

We denote sample correlations between ξ_i and η_i by s_i and the sample canonical correlations between the sample canonical variates by r_i. The r_i may also be interpreted as the cosines of the critical angles between the two planes spanned by $x_1, \cdots, x_p$ and $y_1, \cdots, y_q$ respectively, where the x_i and y_j are the sample canonical variates. The distribution of the r_i for each of the two cases mentioned above will be derived in sections 3 and 4 respectively.

3. Distribution of the canonical correlation coefficients. Our starting point is the distribution (2.1). The distribution of the canonical correlations r_i will be obtained by expressing this distribution in terms of the r_i and other variables and integrating over the ranges of the latter. First of all, let us dispose of the lengths of the vectors ξ_i and η_j.

Put $\xi_i = \tau_i w_i$ and $\eta_j = \sigma_j z_j$ where τ_i and σ_j are the unit vectors along ξ_i and η_j respectively, and w_i and z_j are their lengths. Then

$$\prod_{r=1}^{n}d\xi_{ri} = w_i^{n-1}\,dw_i\,dS(\tau_i) \tag{3.1}$$

where $dS(\tau_i)$ is the element of area on the unit sphere in n-space. With an analogous expression for $\prod d\eta_{rj}$ the distribution (2.1) becomes

$$\begin{aligned}\prod_{i=1}^{p}&\left\{\frac{1}{2^{n-2}(1-\rho_i^2)^{n/2}[\Gamma(n/2)]^2}\right.\\ &\left.\cdot\exp\left[-\frac{1}{2(1-\rho_i^2)}(w_i^2+z_i^2-2\rho_i s_i w_i z_i)\right](w_i z_i)^{n-1}\,dw_i\,dz_i\right\}\\ &\times\prod_{j=p+1}^{q}\frac{1}{2^{(n-2)/2}\Gamma(n/2)}\exp[-\tfrac{1}{2}z_j^2]z_j^{n-1}\,dz_j\prod_{i=1}^{p}\frac{\Gamma(n/2)}{2\pi^{n/2}}dS(\tau_i)\\ &\times\prod_{j=1}^{q}\frac{\Gamma(n/2)}{2\pi^{n/2}}dS(\sigma_j),\end{aligned} \tag{3.2}$$

where $s_i = \tau_i'\sigma_i$ (see section 2). The constants have been split up to make the latter factors probability distributions.

The integrals of the factors containing z_j for $j = p + 1, \cdots, q$ are obviously unity. Furthermore, by expanding the factor $\exp[(1 - \rho_i^2)^{-1}\rho_i s_i w_i z_i]$ in a power series and integrating term-by-term with respect to w_i and z_i $(i = 1, \cdots, p)$ we obtain

$$\int_0^\infty \int_0^\infty \frac{1}{2^{n-2}(1-\rho_i^2)^{n/2}[\Gamma(n/2)]^2} \tag{3.3}$$

$$\cdot \exp[-(w_i^2 + z_i^2 - 2\rho_i s_i w_i z_i)/2(1-\rho_i^2)](w_i z_i)^{n-1}\, dw_i\, dz_i$$

$$= (1-\rho_i^2)^{n/2} {}_2F_1(n/2, n/2; 1/2; \rho_i^2 s_i^2) + \text{an odd function of } s_i,$$

where ${}_2F_1$ is the Gaussian hypergeometric function. Later on, we shall see that the odd function of s_i vanishes in the subsequent integrations.

The next step is to express the unit column vectors τ_i and σ_j in terms of the canonical correlations r_i and the vectors x_i and y_j which determine these correlations. Let $\mathfrak{p}$ and $\mathfrak{q}$ be the p-plane and the q-plane spanned by the vectors τ_i and σ_j in n-space. Then $\mathfrak{p}$ and $\mathfrak{q}$ determine almost certainly (i.e. with probability 1) the orthonormal vectors x_i and y_i $(i = 1, \cdots, p)$ which make the critical angles between the planes, i.e. such that $x_i'y_i = r_i$, $x_i'y_j = 0$ $(i \neq j)$. Let further vectors $y_{p+1}, \cdots, y_q$ be defined as functions of $\mathfrak{p}$ and $\mathfrak{q}$ to complete an orthonormal set spanning $\mathfrak{q}$. T, Σ, X, Y will denote the matrices composed of the column vectors τ_i, σ_j, x_i, y_j, respectively. It follows that $X'X = I_p$, $Y'Y = I_q$ and $X'Y = [R \vdots 0]$ where R is the diagonal matrix with the r_i down the main diagonal. Furthermore we may write

$$T = XA, \qquad \Sigma = YB \tag{3.4}$$

where A is a $p \times p$ and B is a $q \times q$ matrix. Then $T'T = A'A$ and $\Sigma'\Sigma = B'B$. The matrices A and B are subject only to the restriction that all their columns α_i and β_j are of unit length.

We now substitute for $\prod dS(\tau_i)$ and $\prod dS(\sigma_j)$ in (3.2), using the transformations (3.4). To avoid interrupting the continuity of the argument we shall, for the moment, only give the results of the substitution, and defer the proof until section 5. We have then from (5.4)

$$\prod_{i=1}^{p} dS(\tau_i) = |A'A|^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i)\, d\mathfrak{p} + {}^*(dX)\, d\mathfrak{p} \tag{3.5}$$

where $dS(\alpha_i)$ is the element of area on the unit sphere in p-space and $d\mathfrak{p}$ is the differential form representing the invariant measure on the Grassmann manifold of p-planes in n-space. The symbol ${}^*(dX)$ stands for certain differentials involving the elements of X, which, when subsequently multiplied by other differentials, will vanish. Similarly

$$\prod_{j=1}^{q} dS(\sigma_j) = |B'B|^{(n-q)/2} \prod_{j=1}^{q} dS(\beta_j)\, d\mathfrak{q} + {}^*(dY)\, d\mathfrak{q} \tag{3.6}$$

where $dS(\beta_j)$ is the element of area on the unit sphere in q-space. Multiplying (3.5) and (3.6) we obtain

$$(3.7)\quad \prod_{i=1}^{p} dS(\tau_i) \prod_{j=1}^{q} dS(\sigma_j) = |A'A|^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i)\, |B'B|^{(n-q)/2} \prod_{j=1}^{q} dS(\beta_j)\, d\mathfrak{p}\, d\mathfrak{q}.$$

The terms containing $*(dX)$ and $*(dY)$ vanish when multiplied by $d\mathfrak{p}\, d\mathfrak{q}$, since $d\mathfrak{p}\, d\mathfrak{q}$ is of maximum degree in $\mathfrak{p}$ and $\mathfrak{q}$ and X and Y are functions of $\mathfrak{p}$ and $\mathfrak{q}$.

The differential form $d\mathfrak{p}\, d\mathfrak{q}$ may now be expressed in terms of the r_i and other variables. Integration with respect to these latter variables yields

$$K_p K_q \phi(r_i \mid \rho_i = 0)$$

where K_p and K_q are the normalising constants of the differential forms $d\mathfrak{p}$ and $d\mathfrak{q}$ respectively, and $\phi(r_i \mid \rho_i = 0)$ is the null distribution of the r_i (see James [6]):

$$\phi(r_i \mid \rho_i = 0) = C \prod_{i=1}^{p} \{(r_i^2)^{(q-p-1)/2}(1 - r_i^2)^{(n-q-p-1)/2}\} \prod_{i<j} (r_i^2 - r_j^2) \prod_{i=1}^{p} dr_i^2$$

and

$$C = \pi^{p/2} \prod_{i=0}^{p-1} \left\{\Gamma\left(\frac{n-i}{2}\right) \Big/ \Gamma\left(\frac{p-i}{2}\right) \Gamma\left(\frac{q-i}{2}\right) \Gamma\left(\frac{n-q-i}{2}\right)\right\}.$$

This distribution was first derived by Fisher [3], Hsu [5] and Roy [8]. The values of K_p and K_q are given by

$$K_v = \prod_{i=1}^{v} \frac{G(n-i+1)}{G(i)}, \qquad G(i) = \frac{2\pi^{i/2}}{\Gamma(i/2)}, \qquad v = p, q.$$

After this integration, the right hand side of (3.7) becomes

$$(3.8)\quad K_p\, |A'A|^{(n-p)/2} \prod dS(\alpha_i) K_q\, |B'B|^{(n-q)/2} \prod dS(\beta_j) \cdot \phi(r_i \mid \rho_i = 0),$$

showing that A, B and the r_i are independently distributed.

Substituting (3.3) and (3.8) in (3.2), we may write the distribution of the r_i as

$$(3.9)\quad \int_A \int_B \prod_{i=1}^{p} \{(1-\rho_i^2)^{n/2}\, {}_2F_1(n/2, n/2; 1/2; \rho_i^2 s_i^2)\} k_p\, |A'A|^{(n-p)/2} \cdot \prod_{i=1}^{p} dS(\alpha_i) k_q\, |B'B|^{(n-q)/2} \prod_{j=1}^{q} dS(\beta_j) \phi(r_i \mid \rho_i = 0),$$

together with the relation

$$(3.10)\quad s_i = \tau_i' \sigma_i = \alpha_{1i}\beta_{1i} r_1 + \alpha_{2i}\beta_{2i} r_2 + \cdots + \alpha_{pi}\beta_{pi} r_p .$$

The normalising constants k_p and k_q for the distribution of A and B are given by

$$(3.11)\quad k_v = \prod_{i=1}^{v} \frac{G(n-i+1)}{G(n)G(i)}, \qquad v = p, q.$$

In view of equation (3.10) we may now identify our distribution (3.9) with Bartlett's distribution, [1], equations (8) and (10).

If the hypergeometric functions are expanded as power series and multiplied together, the function multiplying $\phi(r_i \mid \rho_i = 0)$ is seen to be a multiple power series in the ρ_i whose coefficients depend on the expectations of monomials in the s_i with respect to the distribution

$$k_p \mid A'A \mid^{(n-p)/2} dS(\alpha_1) \cdots dS(\alpha_p) \tag{3.12}$$

of A and a similar distribution of B.

So far we have ignored the odd function of s_i appearing in the integral (3.3). However, any odd function $f(s_i)$ of s_i will have zero expectation. In fact, putting $-\alpha_i$ instead of α_i does not alter the distribution (3.12) of A, but changes s_i to $-s_i$ in view of (3.10). Therefore,

$$E[f(s_i)] = E[f(-s_i)] = E[-f(s_i)] = -E[f(s_i)]$$

and so $E[f(s_i)] = 0$. It is sufficient, therefore, to compute only moments of the form $\mu(t_1, t_2, \cdots, t_p) = E\{(s_1^2)^{t_1}(s_2^2)^{t_2} \cdots (s_p^2)^{t_p}\}$ where the expectations are taken with respect to the distributions of A and B and the r_i are held fixed. Furthermore, if we substitute in (3.9) for s_i using (3.10), the calculations are reduced to finding the moments of the α_{ij} and β_{ij}, two independent sets of variates.

Theoretically these moments could be found directly from the distributions of A and B. However, as Bartlett pointed out, this method is too difficult algebraically to be of much use, except in the case of only one non-zero ρ_i. Bartlett indicated a method whereby moments of the form $\mu(t_1, t_2)$ could be calculated, and also calculated $\mu(1, 1, 1)$ by employing various relations connecting the α-moments (see section 10). Again, both of these methods led to awkward algebra and had to be abandoned for moments of higher order, though Bartlett was able to compute $\mu(1, 1)$ $\mu(2, 1)$, $\mu(2, 2)$ and $\mu(3, 1)$. In part B of this paper we shall present a method enabling moments of any order to be computed, and shall complete the tabulation of moments up to the fourth order with $\mu(2, 1, 1)$ and $\mu(1, 1, 1, 1)$.

4. The non-central means case. Let $\mathfrak{p}$ be the random plane spanned by the vectors $\xi_1, \cdots, \xi_p$ and $\mathfrak{q}$ the fixed plane spanned by $\eta_1, \eta_2, \cdots, \eta_q$. As we saw in section 2, we may assume that the $\xi_1, \cdots, \xi_p$ are independently distributed and their components ξ_{vi} have the distribution

$$\prod_{i=1}^{p} (2\pi)^{-n/2} \exp\left[-(\xi_i' \xi_i - 2\xi_i' \eta_i + \eta_i' \eta_i)/2\right] \prod_{v=1}^{n} d\xi_{vi}. \tag{4.1}$$

Furthermore, the η_j $(j = 1, \cdots, q)$ may be taken as vectors lying along the first q co-ordinate axes in the sample space and thus having only one non-zero component each, say μ_j in the jth position.

Putting $\xi_i = \tau_i w_i$ as before, (4.1) becomes

$$\prod_{i=1}^{p} \frac{1}{2^{(n-2)/2}\Gamma(n/2)} \exp\left[-(w_i^2 - 2\mu_i s_i w_i + \mu_i^2)/2\right] w_i^{n-1}\, dw_i \tag{4.2}$$
$$\times \prod_{i=1}^{p} \frac{\Gamma(n/2)}{2\pi^{n/2}}\, dS(\tau_i),$$

where $s_i = \tau_{ii}$. The integral with respect to w_i of the ith factor in the first product of (4.2) is ${}_1F_1(n/2; 1/2; \mu_i^2 s_i^2/2)e^{-\mu_i^2/2}$ + an odd function of s_i. This odd function will again vanish in subsequent integrations and may be ignored from now on.

Let X be the $n \times p$ matrix whose columns are the orthonormal vectors $x_1, x_2, \cdots, x_p$ spanning $\mathfrak{p}$ and which make the critical angles with $\mathfrak{q}$. The τ_i may be expressed as linear combinations of the x_i by putting

$$T = XA. \tag{4.3}$$

Since $X'X = I_p$ we have $T'T = A'A$. From section 5, (5.1), it follows that

$$\prod_{i=1}^{p} dS(\tau_i) = |A'A|^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i)\, d\mathfrak{p}, \tag{4.4}$$

the differential form $*(dX)\, d\mathfrak{p}$ vanishing since X and $\mathfrak{p}$ are functions of each other.

To express $\mathfrak{p}$ in terms of the r_i, we partition X as follows:

$$X = \begin{bmatrix} Y \\ \cdots \\ Z \end{bmatrix} \tag{4.5}$$

where Y is a $q \times p$ matrix and Z is an $(n - q) \times p$ matrix. The vector $\begin{bmatrix} y_i \\ \cdots \\ 0 \end{bmatrix}$ in $\mathfrak{q}$ makes the ith critical angle with x_i in $\mathfrak{p}$. Let β_i and $\gamma_i (i = 1, \cdots, p)$ be the unit vectors along y_i and z_i, then according to [6], equation (7.10),

$$y_i = \beta_i r_i, \qquad z_i = \gamma_i \sqrt{1 - r_i^2} \tag{4.6}$$

and

$$d\mathfrak{p} = K_p \frac{1}{\prod_{i=1}^{p} G(q - i + 1)} \tag{4.7}$$
$$\cdot\, dV(\beta) \frac{1}{\prod_{i=1}^{p} G(n - q - i + 1)}\, dV(\gamma)\phi(r_i \mid \rho_i = 0)$$

where K_p and $G(i)$ are defined in section 3, and $dV(\beta)$ and $dV(\gamma)$ are the invariant measures on the Stiefel manifolds of p-frames $(\beta_1, \cdots, \beta_p)$ in q-space and p-frames $(\gamma_1, \cdots, \gamma_p)$ in $(n - q)$-space. The constant has been split up to nor-

malise these invariant measures. If we choose $q - p$ orthonormal vectors β_{p+1}, $\cdots$, β_q orthogonal to $\beta_1, \cdots, \beta_p$ we may express $dV(\beta)$ as

(4.8) $$dV(\beta) = \prod_{i<j}^{p} \beta_j' \, d\beta_i \prod_{j=p+1}^{q} \prod_{i=1}^{p} \beta_j' \, d\beta_i .$$

Also,

$$s_i = \tau_{ii} = \sum_{j=1}^{p} x_{ij}\alpha_{ji} = \sum_{j=1}^{p} \beta_{ij} r_j \alpha_{ji}$$

If we please, we may replace β_{ij} by β_{ji} since they have the same distribution. Integrating (4.7) with respect to γ, substituting in (4.4) and then in (4.2), we obtain the distribution of the r_i as

(4.9) $$\int_A \int_B \prod_{i=1}^{p} {}_1F_1(n/2; 1/2; \tfrac{1}{2}\mu_i^2 s_i^2) e^{-\mu_i^2/2} k_p \, | A'A |^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i) \cdot \frac{1}{\prod_{i=1}^{p} G(q - i + 1)} \beta_j' \, d\beta_i \prod_{j=p+1}^{q} \prod_{i=1}^{p} \beta_j' \, d\beta_i \phi(r_i \mid \rho_i = 0($$

where

(4.10) $$s_i = \alpha_{1i}\beta_{1i}r_1 + \cdots + \alpha_{pi}\beta_{pi}r_p .$$

We notice that the distribution of A is identical with its distribution in the previous case, but now the distribution of B is the invariant distribution on a Stiefel manifold and is independent of n. However, A and B are still independent.

5. Distribution of the co-ordinates of random vectors in a random plane. In relation to the rest of the paper, the purpose of this section is to derive equation (3.5) and a result at the end of section 7. However, the results will be more interesting and intelligible if discussed in terms of probabilities.

$\tau_1, \cdots, \tau_p$ are invariantly distributed unit vectors in n-space, which we write as the columns of an $n \times p$ matrix T. $\mathfrak{p}$ is the plane spanned by the τ_i. We define in $\mathfrak{p}$ a reference set of orthonormal vectors, which we write as the columns of an $n \times p$ matrix X. Thus X is a function of $\mathfrak{p}$ and

(5.1) $$X'X = I_p .$$

Let the column α_i of the $p \times p$ matrix A be the co-ordinates of τ_i relative to the reference set X:

(5.2) $$T = XA.$$

We shall show that $\mathfrak{p}$ is invariantly distributed and that A is independently distributed with density proportional to

(5.3) $$| A'A |^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i).$$

These results are implicit in Bartlett [1] They follow from the lemma which we shall now state and prove. For the application in section 3 we shall have to generalise the situation slightly to include the case when the reference set X is not necessarily a function of $\mathfrak{p}$ alone.

LEMMA. *If T is an $n \times p$ matrix whose columns τ_i are unit vectors, and X and A are $n \times p$ and $p \times p$ matrices satisfying (5.1) and (5.2), then*

$$\prod_{i=1}^{p} dS(\tau_i) = |A'A|^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i)\, d\mathfrak{p} + {}^*(dX)\, d\mathfrak{p} \tag{5.4}$$

where ${}^*(dX)$ is a differential form in X and A, every term of which is of at least first degree in dX. If X is a function of $\mathfrak{p}$ alone, then ${}^*(dX)\, d\mathfrak{p} = 0$

PROOF. Selecting a single column from the matrix equation (5.2) we have

$$\tau_i = X\alpha_i\,. \tag{5.5}$$

Differentiating:

$$d\tau_i = dX\alpha_i + X\, d\alpha_i\,. \tag{5.6}$$

As the differential form for $dS(\alpha_i)$ will be required, we introduce $p - 1$ orthonormal column vectors in p-space orthogonal to α_i. Let C_i be the $p \times p - 1$ matrix with them as columns. Then $dS(\alpha_i)$ is the alternating product of the elements in the vector $C_i'\, d\alpha_i$.

The differential form for $dS(\tau_i)$ requires $n - 1$ orthonormal vectors orthogonal to τ_i. The columns of the matrix XC_i provide $p - 1$ of them, since $C_i'X'\tau_i = C_i'X'X\alpha_i = C_i'\alpha_i = 0$. Choose the remaining $n - p$ orthonormal vectors orthogonal to the plane $\mathfrak{p}$ and let them be columns of an $n \times (n - p)$ matrix B, which is to be a function merely of p.

Premultiply (5.6) by the transpose of the partitioned matrix $[XC_i \,\vdots\, B]$

$$\begin{bmatrix} C_i' X\, d\tau_i \\ B'\, d\tau_i \end{bmatrix} = \begin{bmatrix} C_i' X'\, dX\alpha_i + C_i'\, d\alpha_i \\ B'\, dX\alpha_i \end{bmatrix}. \tag{5.7}$$

Then, the alternating product of the differentials of the vector on the left is $dS(\tau_i)$ and hence the product of all of these for $i = 1, \cdots, p$ is the density on the left-hand side of (5.4).

The alternating product of all the differentials in the right-hand side of (5.7) for $i = 1, \cdots, p$ will give the density in the new co-ordinates. Let us deal with the vector differentials $B'\, dX\alpha_i$ first. These p vector differentials, corresponding to $i = 1, \cdots, p$, comprise the columns of the matrix $B'\, dXA$, of whose elements we therefore want the alternating product. The alternating product of the elements of a *row* of this matrix is $|A|$ times the product of the row of the elements of $B'\, dX$. There being $n - p$ rows in $B'\, dXA$, the alternating product of all its elements is then $|A|^{n-p} \prod_j \prod_i b_j'\, dx_i$. The differential form

$$\prod_j \prod_i b_j'\, dx_i$$

is the invariant measure, $d\mathfrak{p}$, on the Grassmann manifold, i.e. the uniform distribution of a p-plane in n-space (see [6]).

As the differential forms represent probability densities and must therefore be positive, we replace $| A |$ by its modulus $| A'A |^{1/2}$.

The product of the elements of the vector $C_i' \, d\alpha_i$ is $dS(\alpha_i)$. All the products involving an element of $C_i'X' \, dX\alpha_i$ we lump together in the symbol $* \, dX$. Collecting all factors we obtain (5.4). Q.E.D.

We conclude with a result that we shall need in section 7. From (5.1) and (5.2) we have $T'T = A'A$. Hence, if A has the distribution (5.3) then the moments of $A'A$ are the same as the moments of $T'T$ where T has the distribution $\prod dS(\tau_i)$.

Part B

6. Introduction. In this part of the paper we shall be concerned with the problem of calculating the conditional moments

$$\mu(t_1, t_2, \cdots, t_p) = E[(s_1^2)^{t_1}(s_2^2)^{t_2} \cdots (s_p^2)^{t_p}]$$

required for the expansion of the distribution of the canonical correlations r_i.

Recalling the results of sections 3 and 4, we saw that the expectations of monomials in the s_i^2 could be replaced by the expectations of monomials $m(A, B)$ in $\alpha_{ij}\beta_{ij}$ in view of the relation

$$s_i = \alpha_{1i}\beta_{1i}r_1 + \cdots + \alpha_{pi}\beta_{pi}r_p . \tag{6.1}$$

Furthermore, since $A = (\alpha_{ij})$ is distributed independently of $B = (\beta_{ij})$,

$$E[m(A, B)] = E[m(A)] \, E[m(B)]$$

where $m(A)$ and $m(B)$ are monomials in the elements of A and B respectively. Considering case (a) for the moment, we saw that the distributions of A and B were

$$k_p \, | A'A |^{(n-p)/2} \, dS(\alpha_1) \cdots dS(\alpha_p), \tag{6.2}$$

and

$$k_q \, | B'B |^{(n-q)/2} \, dS(\beta_1) \cdots dS(\beta_q)$$

respectively. Consequently, $E[m(B)]$ may be obtained from $E[m(A)]$ by simply replacing p with q.

In case (b), though the distribution of A is still given by (6.2), the distribution of B is given by (4.8), the invariant distribution on the Stiefel manifold of p-frames in q-space. We notice, however, that if we let $n \to \infty$ in case (a), then the set of random vectors $(\beta_1, \cdots, \beta_p)$ becomes a rigid p-frame, and this, of course, is exactly the situation in case (b). Hence the β-moments may be obtained from those in case (a) by letting $n \to \infty$. To summarise, then, it is sufficient to compute only the moments of the distribution (6.2).

To compute these moments by direct integration is obviously going to lead

to involved algebra. However, by first averaging the monomials $m(A)$ over the orthogonal group we can considerably simplify the problem. Before proceeding further we shall briefly discuss this important process.

7. Average over the orthogonal group. The process $\mathfrak{M}$ of averaging over a group is a linear process whereby a function, defined on a space on which a group of transformations acts, is changed into a function invariant under the group. In particular, we consider the group $\mathfrak{H}$ of all orthogonal matrices H, and a matrix $A = (\alpha_{ij})$ which is transformed by the elements of $\mathfrak{H}$:

$$A \to HA \tag{7.1}$$

If $f(A)$ is a function of the elements of A, then

$$\mathfrak{M}f(A) = \int_{\mathfrak{H}} f(H^{-1}A)\, dV(H)$$

is a function invariant under the transformations (7.1). $V(H)$ is the invariant measure on the orthogonal group, normalised so that $V(\mathfrak{H}) = 1$. $\mathfrak{M}f$ is called the *average or mean value* of the function over the group. (This definition of "mean value" should not be confused with the usual statistical definition.) Since $\mathfrak{M}f$ is invariant under the orthogonal group, it must be expressible as a function of the basic invariants $\alpha_i'\alpha_j$ (see Weyl [9], pp. 52–6).

We wish to calculate the expectations of monomials $m(A)$ in the elements of A. Since the distribution (6.2) is invariant under the transformations (7.1), $E[m(A)] = E[m(H^{-1}A)]$, and hence it follows that

$$E[m(A)] = \int E[m(A)]\, dV(H) = \int E[m(H^{-1}A)]\, dV(H)$$

$$= E\int m(H^{-1}A)\, dV(H) = E[\mathfrak{M}m(A)].$$

In section 8 we shall show how to calculate $\mathfrak{M}m(A)$.

However, assuming for the moment that this has been done, we see that the problem has been reduced to the evaluation of the expectations of certain invariant functions $\phi(A'A)$, say. At this point it should be noted that the problem of the β-moments in case (b) has been completely solved. For, if we let $n \to \infty$, then $B'B = I$ with probability 1, and hence $E[m(B)] = \phi(I)$. $E[m(B)]$ can be then evaluated by the method given in James [7], pp. 371–5. However, since we require the β-moments for case (a), we may as well compute the α for case (b) by letting $n \to \infty$ in the former moments, as indicated in section 6.

For the α-moments (and the β-moments for case (a)), we still have to evaluate the expectations of the invariant functions. In section 5 we have shown that the $\alpha_i'\alpha_j$ have the same distribution as quantities $r_i'r_j$ where $r_1, \cdots, r_p$ are independently uniformly distributed unit vectors in n-space. Finally, then, there remains the calculation of the moments of the $r_i'r_j$. This will be accomplished in section 9.

8. Calculation of $\mathfrak{M}m(A)$. In section 7 it was shown that

$$E[m(A)] = E[\mathfrak{M}m(A)] = E[\phi(A'A)].$$

In this section we shall show how to evaluate $\mathfrak{M}m(A)$.

Let

$$m(A) = \alpha_{i_1j_1}^{k_1}\alpha_{i_2j_2}^{k_2}\cdots \tag{8.1}$$

denote a monomial in the α_{ij}. Then if C is an arbitrary $p \times p$ matrix, the expansion of the function exp (tr $C''A$) contains every monomial (8.1) multiplied by the same monomial $m(C)$ in the corresponding elements of C, and divided by $k_1 ! \, k_2 ! \cdots$. James [7] has shown that $\mathfrak{M}$ exp (tr $C'A$) can be expanded as a multiple power series in the elementary symmetric functions $z_1, z_2, \cdots, z_p$ of the latent roots of $C'CA'A$. Thus, if $\lambda_1, \cdots, \lambda_p$ are the latent roots of $C'CA'A$, then

$$z_1 = \sum \lambda_i = \text{tr } C'CA'A,$$

$$z_2 = \sum_{i<j} \lambda_i\lambda_j = \text{sum of principal 2nd order minors of } C'CA'A, \text{ etc.,}$$

and

$$\begin{aligned}\mathfrak{M}\exp(\text{tr } C'A) &= 1 + \frac{z_1}{2p} + \frac{z_1^2}{8p(p+2)} + \frac{z_2}{2p(p+2)(p-1)} \\ &+ \frac{z_1^3}{8\cdot 3!\,p(p+2)(p+4)} + \frac{z_1 z_2}{4p(p+2)(p+4)(p-1)} \\ &+ \frac{z_3}{p(p+2)(p+4)(p-1)(p-2)} + \frac{z_1^4}{2^4 4!\,p(p+2)(p+4)(p+6)} \\ &+ \frac{z_1^2 z_2}{16p(p+2)(p+4)(p+6)(p-1)} \\ &+ \frac{z_2^2}{8p(p+2)(p+4)(p+6)(p-1)(p+1)} \\ &+ \frac{(p+2)z_1z_3}{2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)} \\ &+ \frac{(5p+6)z_4}{2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)} + \cdots.\end{aligned} \tag{8.2}$$

Hence, $\mathfrak{M}m(A)$ can be found by equating the coefficients of $m(C)$ on both sides of (8.2).

If we write $A'A$ in the form

$$\begin{bmatrix} 1 & \alpha_1'\alpha_2 & \alpha_1'\alpha_3 & \cdots & \alpha_1'\alpha_p \\ \alpha_1'\alpha_2 & 1 & \alpha_2'\alpha_3 & \cdots & \cdot \\ \alpha_1'\alpha_3 & \alpha_2'\alpha_3 & 1 & & \cdot \\ \vdots & \vdots & \vdots & & \vdots \\ \alpha_1'\alpha_p & \alpha_2'\alpha_p & & & 1 \end{bmatrix} \tag{8.3}$$

we see that $\mathfrak{M}m(A)$ will be a linear combination of monomials in the invariants $\alpha_i'\alpha_j$. The expansion (8.2) is sufficient to compute all conditional moments up to order 4. If higher moments are required, further terms can be added to (8.2) by the use of recurrence relations derived from the differential equations given in James [7].

9. Calculation of the moments of the invariants. We are given that $\tau_1, \tau_2, \cdots, \tau_p$ are independently uniformly distributed column vectors in n-space, and we require the expectations of monomials in $\tau_i'\tau_j$. If a monomial in $\tau_i'\tau_j$ were expanded as a sum of monomials in the τ_{ij}, the expectations of each of these could be calculated and summed. However, the expansions would become very complicated. They can be avoided by the following method, which is an extension of an idea due to Bartlett [1] p 13

Let $c_1, c_2, \cdots, c_p$ be the unit vectors along the first p coordinate axes. Then the joint distribution of $\tau_1, \cdots, \tau_p$ is the same as that of $A_1c_1, A_2c_2, \cdots, A_pc_p$, where the A_i are random orthogonal matrices independently and invariantly distributed (see James [6]) Furthermore, the invariant functions will not be altered if they are calculated from the vectors $c_1, A_1'A_2c_2, \cdots, A_1'A_pc_p$. These vectors have the same distribution as $c_1, A_2c_2, \cdots, A_pc_p$ since $A_1'A_2, \cdots, A_1'A_p$ are still independently invariantly distributed Again, if $A_2 = (a_{ij})$, say, the invariant functions will not be altered if we replace the vectors by

$$c_1, B_2'A_2c_2, \cdots, B_2'A_pc_p$$

where

$$B_2 = \begin{bmatrix} 1 & 0 & 0 & \cdot & 0 \\ 0 & a_{22}/b_{22} & * & & \\ 0 & a_{32}/b_{22} & * & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ 0 & a_{n2}/b_{22} & & \cdot & \end{bmatrix},$$

$b_{22}^2 = 1 - a_{12}^2 = a_{22}^2 + \cdots + a_{n2}^2$, and the remaining elements are chosen so that B_2 is orthogonal. Clearly,

$$B_2'A_2c_2 = \begin{bmatrix} a_{12} \\ b_{22} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Since the matrices $B_2'A_3, \cdots, B_2'A_p$ are still independently invariantly distributed we may replace the vectors by

$$c_1, B_2'A_2c_2, A_3c_3, \cdots, A_pc_p.$$

Proceeding in this way we see that we obtain the same values for the expectations of the invariants if we replace $\tau_1, \tau_2, \cdots, \tau_p$ by

$$(9.1) \qquad \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \quad \begin{bmatrix} a_{12} \\ b_{22} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \quad \begin{bmatrix} a_{13} \\ a_{23} \\ b_{33} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \cdots.$$

(To avoid introducing further notation, we have denoted the third column of A_3 by the elements $a_{13}, a_{23}, \cdots, a_{n3}$, those of the fourth column of A_4 by $a_{14}, a_{24}, \cdots, a_{n4}$ etc. Then $b_{33}^2 = 1 - a_{13}^2 - a_{23}^2$, $b_{44}^2 = 1 - a_{14}^2 - a_{24}^2 - a_{34}^2$, etc.)

EXAMPLE 1. As an example let us evaluate

$$E[(\alpha_1'\alpha_2)(\alpha_2'\alpha_3)(\alpha_3'\alpha_4)(\alpha_1'\alpha_4)] = E[(\tau_1'\tau_2)(\tau_2'\tau_3)(\tau_3'\tau_4)(\tau_1'\tau_4)].$$

Substituting from (9.1), this expectation is equal to

$$(9.2) \qquad E[a_{12}(a_{12}a_{13} + b_{22}a_{23})(a_{13}a_{14} + a_{23}a_{24} + b_{33}a_{34})a_{14}].$$

Now any monomial in the a_{ij}, b_{ii} containing an odd power has zero expectation since the distribution is unaltered if we replace a_{ij} by $-a_{ij}$ or b_{ii} by $-b_{ii}$. Hence, (9.2) reduces to $E(a_{12}^2a_{13}^2a_{14}^2)$. a_2, a_3 and a_4 are independently uniformly distributed unit vectors, and hence $E(a_{12}^2) = E(a_{13}^2) = E(a_{14}^2) = 1/n$. Therefore,

$$E[(\alpha_1'\alpha_2)(\alpha_2'\alpha_3)(\alpha_3'\alpha_4)(\alpha_1'\alpha_4)] = 1/n^3.$$

EXAMPLE 2. $E(\Delta)$ where $\Delta = | A'A |$.
Put

$$C = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots \\ 0 & b_{22} & a_{33} & \cdots \\ 0 & 0 & b_{33} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}.$$

Then $\Delta = | C'C | = | C |^2$, and

$$E(\Delta) = E(1 \cdot b_{22}^2 \cdot b_{33}^2 \cdots b_{pp}^2)$$

$$= 1 \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-p+1}{n},$$

since $E(b_{22}^2) = 1 - E(a_{12}^2) = 1 - 1/n$, etc.

10. Example of the calculation of the conditional moments. Following Bartlett, we introduce the notation[2]

$$(10.1)\qquad \begin{aligned} E(\alpha_{11}^2\alpha_{12}^2)E(\beta_{11}^2\beta_{12}^2) &= \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \\ E(\alpha_{11}^2\alpha_{22}^2)E(\beta_{11}^2\beta_{22}^2) &= \begin{pmatrix} 2 & \cdot \\ \cdot & 2 \end{pmatrix}, \quad \text{etc.} \end{aligned}$$

From equation (6.1) it is seen that the conditional moments can be expressed as linear combinations of "arrays" similar to those in (10.1). As we saw in section 6, it is sufficient to calculate the α-moments only.

To illustrate the method let us calculate the α-moment or "half-factor" corresponding to

$$\begin{pmatrix} 1 & 1 & \cdot \\ & 1 & 1 \\ 1 & & 1 \end{pmatrix},$$

i e. $E(\alpha_{11}\alpha_{13}\alpha_{21}\alpha_{22}\alpha_{32}\alpha_{33})$.

The first step is to calculate $\mathfrak{M}m(A)$. Now,

$$C'CA'A = \begin{bmatrix} * & . \; c_{21}c_{22} + \cdots & . \; c_{11}c_{13} + \cdots & . \; \cdots \\ c_{21}c_{22} + \cdots . & * & . \; c_{32}c_{33} + & \cdot \; \cdot \; \cdots \\ c_{11}c_{13} + \cdots . & c_{32}c_{33} + \cdot & * & \cdot \; \cdot \\ \cdots\cdots\cdots\cdots & & \cdot & \cdot\;\cdot\;\cdot \end{bmatrix}$$

$$\begin{bmatrix} 1 & \alpha_1'\alpha_2 & \alpha_1'\alpha_3 & \cdots \\ \alpha_1'\alpha_2 & 1 & \alpha_2'\alpha_3 & \cdots \\ \alpha_1'\alpha_3 & \alpha_2'\alpha_3 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

All remaining terms in $C'C$ may be neglected as they will not contribute to $m(C)$ in the expansion (8.2).

$$\therefore z_1^3 = 48(\alpha_1'\alpha_2)(\alpha_2'\alpha_3)(\alpha_1'\alpha_3)m(C) + \cdots,$$

$$z_1z_2 = 4\{3(\alpha_1'\alpha_2)(\alpha_2'\alpha_3)(\alpha_1'\alpha_3) - (\alpha_1'\alpha_2)^2 - (\alpha_2'\alpha_3)^2 - (\alpha_1'\alpha_3)^2\}m(C) + \cdots.$$

$$z_3 = 2m(C)\begin{vmatrix} 1 & \alpha_1'\alpha_2 & \alpha_1'\alpha_3 \\ \alpha_1'\alpha_2 & 1 & \alpha_2'\alpha_3 \\ \alpha_1'\alpha_3 & \alpha_2'\alpha_3 & 1 \end{vmatrix} + \cdots.$$

Hence, after calculating the expectations of the invariant functions by the method of section 9, we have

$$E(z_1^3) = \frac{48}{n^3}m(C) + \cdots$$

$$E(z_1z_2) = -\frac{12(n-1)}{n^3}m(C) + \cdots$$

$$E(z_3) = \frac{2(n-1)(n-2)}{n^3}m(C) + \cdots$$

---

[2] Actually our notation differs slightly from Barlett's. Whereas [illegible] of rows vectors, we have worked in terms of column vectors, [illegible] responds to our α_{ji}.

Substituting in (8.2) and equating the coefficients of $m(C)$, we obtain after simplification

$$E(\alpha_{11}\,\alpha_{13}\,\alpha_{21}\,\alpha_{22}\,\alpha_{32}\,\alpha_{33}) = \frac{(n-p)(2n-p)}{n^2p(p+2)(p+4)(p-1)(p-2)},$$

which agrees with the value tabulated by Bartlett.

Any other α-moment can be calculated in a similar fashion. In particular, the moments tabulated by Bartlett were checked and the various terms contained in $\mu(2, 1, 1)$ and $\mu(1, 1, 1, 1)$ have been calculated and included in the appendix. Actually, only the α-moments have been tabulated. The complete value for case (a) may be obtained by multiplying the α-moment by a similar value with q replacing p. The complete value for case (b) is obtained by taking the previous value and letting $n \to \infty$ in the second half.

Incidentally, the α-moments may be checked by an independent method. For example, consider the monomial $\alpha_{11}^4\alpha_{12}^2$. If we multiply it by $\alpha_3'\alpha_3$, which is identically unity, then $E[\alpha_{11}^4\alpha_{12}^2(\alpha_3'\alpha_3)] = E[\alpha_{11}^4\alpha_{12}^2]$. But expanding the term on the left-hand side, we get

$$E[\alpha_{11}^4\alpha_{12}^2] = E[\alpha_{11}^4\alpha_{12}^2\alpha_{13}^2] + E[\alpha_{11}^4\alpha_{12}^2\alpha_{23}^2] + E[\alpha_{11}^4\alpha_{12}^2\alpha_{33}^2] + \cdots,$$

and therefore

$$\begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}4\\2\\2\end{pmatrix} + (p-1)\begin{pmatrix}4 & \cdot\\2 & \cdot\\\cdot & 2\end{pmatrix}.$$

Similarly, by expanding $(\alpha_1'\alpha_2)^2(\alpha_1'\alpha_3)^2$, whose expectation $= 1/n^2$, we have

$$p\begin{pmatrix}4\\2\\2\end{pmatrix} + p(p-1)\begin{pmatrix}2 & 2\\2 & \cdot\\\cdot & 2\end{pmatrix} + 4p(p-1)\begin{pmatrix}3 & 1\\1 & 1\\2 & \cdot\end{pmatrix} + 2p(p-1)\begin{pmatrix}2 & 2\\1 & 1\\1 & 1\end{pmatrix}$$
$$+ 2p(p-1)(p-2)\begin{pmatrix}2 & 1 & 1\\\cdot & 1 & 1\\2 & \cdot & \cdot\end{pmatrix} + 4p(p-1)(p-2)\begin{pmatrix}2 & 1 & 1\\1 & 1 & \cdot\\1 & \cdot & 1\end{pmatrix}$$
$$+ p(p-1)(p-2)(p-3)\begin{pmatrix}1 & 1 & 1 & 1\\1 & 1 & \cdot & \cdot\\\cdot & \cdot & 1 & 1\end{pmatrix} = 1/n^2.$$

REFERENCES

[1] M. S. Bartlett, "The general canonical correlation distribution," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 1–17.

[2] R. A. Fisher, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc.*, Vol. A121 (1928), pp. 654–673.

[3] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eug.*, Vol. 9 (1939), pp. 238–249.

[4] H. Hotelling, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936) pp. 321–377.

[5] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Ann. Eug.*, Vol. 9 (1939), pp. 250-258
[6] A T. James, "Normal multivariate analysis and the orthogonal group," *Ann Math Stat.*, Vol. 25 (1954), pp. 40-75.
[7] A T James, "A generating function for averages over the orthogonal group," *Proc. Roy. Soc.*, Vol. A229 (1955), pp. 367-375.
[8] S N. Roy, "*p*-statistics or some generalisations in analysis of variance appropriate to multivariate problems," *Sankhya*, Vol. 4 (1939), pp. 381-396.
[9] H. Weyl, *The Classical Groups*, Princeton University Press, 2nd ed , 1946

APPENDIX

$$\begin{pmatrix}4\\2\\2\end{pmatrix} = \frac{3(n+4)(n+6)}{n^2p(p+2)(p+4)(p+6)} \cdot \begin{pmatrix}4 & \cdot\\2 & \cdot\\ \cdot & 2\end{pmatrix} = \frac{3(n+4)(np+5n-6)}{n^2p(p+2)(p+4)(p+6)(p-1)}.$$

$$\begin{pmatrix}2 & 2\\2 & \cdot\\2 & \cdot\end{pmatrix} = \frac{(n+4)(np+3n+2p-6)}{n^2p(p+2)(p+4)(p+6)(p-1)} \cdot \begin{pmatrix}3 & 1\\1 & 1\\2 & \cdot\end{pmatrix} = \frac{-3(n-p)(n+4)}{n^2p(p+2)(p+4)(p+6)(p-1)}.$$

$$\begin{pmatrix}4 & \cdot\\1 & 1\\1 & 1\end{pmatrix} = \frac{-3(n-p)(n+4)}{n^2(p+2)(p+4)(p+6)(p-1)}.$$

$$\begin{pmatrix}4 & \cdot\\ \cdot & 2\\ \cdot & 2\end{pmatrix} = \frac{3\{n^2(p+3)(p+5)+2n(p+1)(p+3)-8(2p+3)\}}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix}2 & 2\\2 & \cdot\\ \cdot & 2\end{pmatrix} = \frac{n^2(p+3)^2+2n(p+1)(2p+3)+4(p^2-4p-6)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix}3 & 1\\1 & 1\\ \cdot & 2\end{pmatrix} = \frac{-3(n-p)(np+3n+2p)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)} \cdot \begin{pmatrix}2 & 2\\1 & 1\\1 & 1\end{pmatrix} = \frac{-(n-p)(np-3n+8p+12)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix}4 & \cdot & \cdot\\ \cdot & 2 & \cdot\\ \cdot & \cdot & 2\end{pmatrix} = \frac{3\{n^2(p^3+8p^2+13p-2)-2n(5p^2+27p+22)+8(5p+6)\}}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix}2 & 2 & \cdot\\2 & \cdot & \cdot\\ \cdot & \cdot & 2\end{pmatrix} = \frac{n^2(p^3+6p^2+3p-6)+2n(p^3-19p-18)-4(3p^2-8p-12)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} \cdot & 2 & 2 \\ 2 & \cdot & \cdot \\ 2 & \cdot & \cdot \end{pmatrix} = \frac{n^2(p+3)(p+5) + 2n(p+1)(p+3) - 8(2p+3)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)}$$

$$\begin{pmatrix} 3 & 1 & \cdot \\ 1 & 1 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} = \frac{-3(n-p)(np^2 + 5np + 2n - 6p - 4)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 4 & \cdot & \cdot \\ \cdot & 1 & 1 \\ \cdot & 1 & 1 \end{pmatrix} = \frac{-3(n-p)(np^2 + 7np + 14n - 8p - 16)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & \cdot & 2 \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \end{pmatrix} = \frac{-(n-p)(np^2 + 3np + 6n + 2p^2 - 6p - 12)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}$$

$$\begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & \cdot \\ 2 & \cdot & \cdot \end{pmatrix} = \frac{-(n-p)(np + 3n + 2p)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & \cdot \\ 1 & \cdot & 1 \end{pmatrix} = \frac{-(n-p)(3p^2 - 4np + p - 6)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & 1 & 1 \\ \cdot & 1 & 1 \\ 2 & \cdot & \cdot \end{pmatrix} = \frac{-(n-p)(np^2 + 3np + 6n + 2p^2 - 6p - 12)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}$$

$$\begin{pmatrix} 3 & 1 & \cdot \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \end{pmatrix} = \frac{3(n-p)(2np + 4n - p^2 - p - 2)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 2 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} = \frac{n^2(p^3 + 8p^2 + 13p - 2) - 2n(5p^2 + 27p + 22) + 8(5p + 6)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}$$

$$\begin{pmatrix} 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & 1 & 1 \end{pmatrix} = \frac{-(n-p)(np^2 + 7np + 14n - 8p - 16)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \end{pmatrix} = \frac{(n-p)(n-p-2)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix} 2 & 1 & 1 & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} = \frac{-(n-p)(np^2 + 5np + 2n - 6p - 4)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot & 2 \\ \cdot & 1 & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \end{pmatrix} = \frac{(n-p)(2np + 4n - p^2 - p - 2)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} = \frac{(n+2)(n+4)(n+6)}{n^3p(p+2)(p+4)(p+6)}. \quad \begin{pmatrix} 2 & \cdot \\ 2 & \cdot \\ 2 & \cdot \\ \cdot & 2 \end{pmatrix} = \frac{(n+2)(n+4)(np + 5n - 6)}{n^3p(p+2)(p+4)(p+6)(p-1)}.$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 2 & \cdot \\ 2 & \cdot \end{pmatrix} = \frac{-(n-p)(n+2)(n+4)}{n^3p(p+2)(p+4)(p+6)(p-1)}.$$

$$\begin{pmatrix} 2 & \cdot \\ 2 & \cdot \\ \cdot & 2 \\ \cdot & 2 \end{pmatrix} = \frac{(n+2)\{n^2(p+3)(p+5) + 2n(p+1)(p+3) - 8(2p+3)\}}{n^3(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix} 2 & \cdot \\ & 2 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{-(n-p)(n+2)(np+3n+2p)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \frac{3(n-p)(n+2)(n-p-2)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix} 2 & \cdot & \cdot \\ 2 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} = \frac{(n+2)\{n^2(p^3+8p^2+13p-2) - 2n(5p^2+27p+22) + 8(5p+6)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & \cdot & \cdot \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} = \frac{-(n-p)(n+2)(np^2+5np+2n-6p-4)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 1 & \cdot & 1 \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ 1 & \cdot & 1 \end{pmatrix} = \frac{(n+2)(n-p)(n-p-2)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \\ 2 & \cdot & \cdot \end{pmatrix} = \frac{(n-p)(n+2)(2np+4n-p^2-p-2)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}$$

$$\begin{pmatrix} [illegible] \end{pmatrix} = \frac{-(n-p)(n+2)(np^2+7np+14n-8p-16)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} [illegible] \end{pmatrix} = \frac{n^3(p^4+7p^3+p^2-35p-6) - 12n^2(p^3+6p^2+3p-6) + 4n(19p^2+79p+51) - 48(5p+6)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & 1 & 1 \end{pmatrix} = \frac{(n-p)\{n^2(p^2+5p+18) - n(p^3+5p^2+18p) + 4(2p^2+3p-6)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} = \frac{(n-p)\{2n^2(p^2+4p) - n(p^3+4p^2+15p+18) + 6(p^2+p+2)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

$$\begin{pmatrix} 2 & \cdot & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} = \frac{-(n-p)\{n^2(p^3+6p^2+3p-6) - 2n(5p^2+21p+18) + 8(5p+6)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & 1 & 1 \\ 1 & \cdot & \cdot & 1 \end{pmatrix} = \frac{-(n-p)\{n^2(5p+6) - n(5p^2+6p) + (p^3+p^2+2p)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

SIGNIFICANCE LEVEL AND POWER[1]

BY E. L. LEHMANN

University of California, Berkeley

1. Summary and introduction. Significance testing, as described in most textbooks, consists in fixing a standard significance level α such as .01 or .05 and rejecting the hypothesis $\theta = \theta_0$ if a suitable statistic Y exceeds C where $P_{\theta_0}\{Y > C\} = \alpha$. Such a procedure controls the probability of false rejection (error of the first kind) at the desired level α but leaves the power of the test and hence the probability of an error of the second kind to the mercy of the experiment. It seems more natural when deciding on a significance level (and this suggestion is certainly not new) to take into account also what power can be achieved with the given experiment. In Section 3 a specific suggestion will be made as to how to balance α against the power β obtainable against the alternatives of interest.

The adoption of this or some similar rule for choosing a significance level has important consequences for the theory of testing composite hypotheses, where nuisance parameters are present. Since the quantity α is then potentially a function of the nuisance parameter ϑ, the classical rule of a fixed significance level leads to the condition that the tests be *exact* or *similar*, that is, that $\alpha(\vartheta)$ equal the preassigned value α for all ϑ. On the other hand, the power β that can be attained against any alternative $\theta = \theta_1$ frequently depends on ϑ. The requirement that $\alpha(\vartheta)$ and $\beta(\vartheta)$ be in a certain balance thus leads to tests which are not similar and hence do not agree with the standard solutions.

To obtain a suitable setting for this discussion, we consider first a minimal complete class of tests for testing the hypothesis $H:\theta \leq \theta_0$ in a multiparameter exponential family (Section 2). The proposed α, β-relation is discussed in Section 3, and in Section 4 is applied to the exponential family. Section 5 gives some illustrations of the theory.

2. A complete class theorem. Many standard testing problems concern an exponential family of distributions, which has probability densities of the form

$$p_{\theta,\vartheta}(x) = C(\theta, \vartheta) \exp\left[\theta U(x) + \sum_{i=1}^{r} \vartheta_i T_i(x)\right] h(x) \tag{1}$$

with respect to a σ-finite measure μ, where θ, U, the ϑ_i and T_i are real-valued and where $\vartheta = (\vartheta_1, \cdots, \vartheta_r)$. In this family, the statistics U and $T = (T_1, \cdots, T_r)$ constitute a set of sufficient statistics for (θ, ϑ).

The problem of testing the hypothesis $H:\theta \leq \theta_0$ against the one-sided al-

Received January 17, 1958; revised July 22, 1958.

[1] This paper was prepared with the partial support of the Office of Naval Research. This paper in whole or in part may be reproduced for any purpose of the United States Government.

ternatives $\theta > \theta_0$ has been treated by many authors (usually in the formulation $\theta = \theta_0$ against $\theta > \theta_0$). The solution of this testing problem according to the Neyman-Pearson theory is the uniformly most powerful unbiased test; this depends only on U and T and is given by the critical function[2]

$$\phi(u, t) = \begin{cases} 1 \text{ if } u > C(t), \\ \gamma(t) \text{ if } u = C(t), \\ 0 \text{ if } u < C(t), \end{cases} \tag{2}$$

where the functions C and γ are determined by the conditions $E_{\theta_0}[\phi(U, T) \mid T = t] = \alpha$ and $E_{\theta_0}[U\phi(U, T) \mid T = t] = \alpha\, E_{\theta_0}[U \mid T = t]$ for all t. The condition of unbiasedness

$$E_{\theta,\vartheta}\phi(U, T) \lesseqqgtr \alpha \quad \text{as} \quad \theta \lesseqqgtr \theta_0,$$

and that of similarity

$$E_{\theta_0,\vartheta}\phi(U, T) = \alpha \quad \text{for all } \vartheta$$

which it implies and which by itself is sufficient to justify the test, are not inherent in the problem but are imposed, at least in part, to facilitate the solution. Before proposing an alternative approach, it is interesting to see how far the problem can be reduced without the introduction of extraneous principles. This can be done by viewing it within the framework of decision theory.

Let d_0 and d_1 denote the decisions of accepting and rejecting the hypothesis H, and denote by $L_i(\theta, \vartheta)$ the loss resulting from decision d_i when (θ, ϑ) are the true parameter values. Then for fixed ϑ, the function $L_0(\theta, \vartheta)$ typically will be zero for $\theta \leqq \theta_0$ and increasing for $\theta \geqq \theta_0$, while $L_1(\theta, \vartheta)$ will be decreasing for $\theta \leqq \theta_0$ and zero for $\theta \geqq \theta_0$. In particular, the difference then satisfies

$$L_1(\theta, \vartheta) - L_0(\theta, \vartheta) \gtreqqless 0 \quad \text{as} \quad \theta \lesseqqgtr \theta_0. \tag{3}$$

The risk function of a test ϕ, which is the expected loss resulting from its use considered as a function of the parameters, is

$$R_\varphi(\theta, \vartheta) = \int \{\varphi(U(x), T(x))L_1(\theta, \vartheta) + [1 - \varphi(U(x), T(x))]L_0(\theta, \vartheta)\}\, p_{\theta,\vartheta}(x)\, d\mu(x). \tag{4}$$

Let $\mathcal{C}$ be the class of all tests satisfying (2) for some functions C and γ. For all loss functions satisfying (3) it was shown by Truax [13] that $\mathcal{C}$ is essentially complete; that is, given any φ there exists $\varphi' \in \mathcal{C}$ such that

$$R_{\varphi'}(\theta, \vartheta) \leqq R_\varphi(\theta, \vartheta) \quad \text{for all } (\theta, \vartheta). \tag{5}$$

We shall now prove that among essentially complete classes, $\mathcal{C}$ is minimal in the sense that if (5) holds for two tests φ, φ' in $\mathcal{C}$, then $\varphi = \varphi'$ a.e. μ.*

[2] See for example [7].

* Recently I learned that this result has been obtained also by D. L. Burkholder. His results are sketched in Abstract 18, *Ann. Math. Stat.*, Vol. 29 (1958), p. 616.

Let φ and φ' belong to $\mathcal{C}$ and let

(6) $$\alpha(\vartheta) = E_{\theta_0,\vartheta}\,\varphi(U, T), \qquad \alpha'(\vartheta) = E_{\theta_0,\vartheta}\,\varphi'(U, T)$$

(i) If the functions α and α' do not agree for all ϑ, suppose without loss of generality that there exists ϑ_0 such that $\alpha(\vartheta_0) < \alpha'(\vartheta_0)$. Since for $\vartheta = \vartheta_0$, the expected values of φ and φ' are continuous functions of θ, there exist $\theta_1 < \theta_0 < \theta_2$ such that

(7) $$E_{\theta,\vartheta_0}\varphi(U, T) < E_{\theta,\vartheta_0}\varphi'(U, T) \qquad \text{for } \theta = \theta_1 \text{ and } \theta = \theta_2 .$$

Then $R_\phi(\theta_1, \vartheta_0) < R_{\phi'}(\theta_1, \vartheta_0)$ and $R_\phi(\theta_2, \vartheta_0) > R_{\phi'}(\theta_2, \vartheta_0)$, and hence neither of the procedures ϕ and ϕ' is uniformly better than the other. (ii) Suppose on the other hand that $\alpha(\vartheta) \equiv \alpha'(\vartheta)$. The standard proof showing a similar test satisfying (2) to be uniformly most powerful similar also shows that a test ϕ_0 satisfying (2) and

(8) $$E_{\theta_0,\vartheta}\,\phi_0(U, T) = \alpha(\vartheta) \qquad \text{for all } \vartheta$$

is uniformly most powerful among all tests satisfying (8). The tests ϕ and ϕ' are therefore both uniformly most powerful within this class and hence

$$E_{\theta,\vartheta}\,\phi(U, T) = E_{\theta,\vartheta}\,\phi'(U, T) \qquad \text{for all } \theta > \theta_0 \text{ and all } \vartheta.$$

Since the family of distributions of the sufficient statistics (U, T) is complete, it follows that $\phi(u, t) = \phi'(u, t)$ a.e., as was to be proved.

3. Significance level and power. It follows from the result of the preceding section that the class $\mathcal{C}$ of tests (2) represents the maximum reduction that can be achieved by comparing only tests of which one has a uniformly better risk function than the other. The selection of a specific test from $\mathcal{C}$, involves two difficulties. It requires the adoption of some principle (Bayes, minimax, etc.) leading to a definite choice;[3] in addition, it requires knowledge of the loss functions L_0 and L_1. An alternative approach, utilizing the fortunate circumstance that the complete class is independent of the actual loss functions (subject only to their satisfying (3)), consists in making the choice by some simple rule of thumb, which does not require (the usually unavailable) knowledge of these losses.

Consider the simplest case of the family (1) with $r = 0$, which involves no nuisance parameters. The family of tests (2) is then a one-parameter family, one test corresponding to each value of

$$\alpha_0 = E_{\theta_0}\phi(X), \qquad 0 \leq \alpha_0 \leq 1.$$

A simple method of choice consists in specifying a value of α_0 and selecting the test corresponding to this value. This need not be a purely formal or arbitrary

[3] Particular proposals of this kind that have been made in the literature include those of Jeffreys [5] involving considerations of *a priori* probabilities, and of Lindley [8] based on his concept of unlikelihood.

procedure since α_0 as the maximum probability of false rejection is of course an important quantity in its own right.

Nevertheless, as was pointed out in Section 1, the above rule appears to neglect too many aspects of the problem. In particular, suppose that the alternatives of primary interest, for which it is important to reject the hypothesis, are those satisfying $\theta \geqq \theta_1$ ($\theta_0 < \theta_1$). Since the power function of any test (2) is increasing in θ, the probability β_1 of rejection when $\theta = \theta_1$ is the minimum power against these alternatives. It seems then reasonable that the choice of test should involve at least β_1 in addition to α_0.

The quantities α_0 and $\alpha_1 = 1 - \beta_1$ are the error probabilities associated with the problem of testing the simple hypothesis $\theta = \theta_0$ against the simple alternative $\theta = \theta_1$. The attainable pairs (α_0, α_1) form a convex set, the lower boundary of which corresponds to the admissible tests (2). This lower boundary is a convex curve S connecting the points (0, 1) and (1, 0), and what is needed is a reasonable way of selecting a point on each such curve. One possible approach to this question is in terms of indifference curves. Suppose that a system of curves could be specified in the (α_0, α_1)-plane such that any two points lying on the same curve are equally desirable, with the curves closer to the origin being more desirable than those further away. The optimum test would then be given by that point of S lying on the indifference curve closest to the origin (Fig. 1).

It seems likely that even this approach is too complex for most applications. To obtain an even simpler formulation, consider once more the rule of fixing the significance level without regard to power. If the level is α, this means restricting attention to the points (α_0, α_1) lying on the vertical line segment $L: \alpha_0 = \alpha$, $0 \leqq \alpha_1 \leqq 1 - \alpha$. The test then corresponds to the point (α_0, α_1), which is the

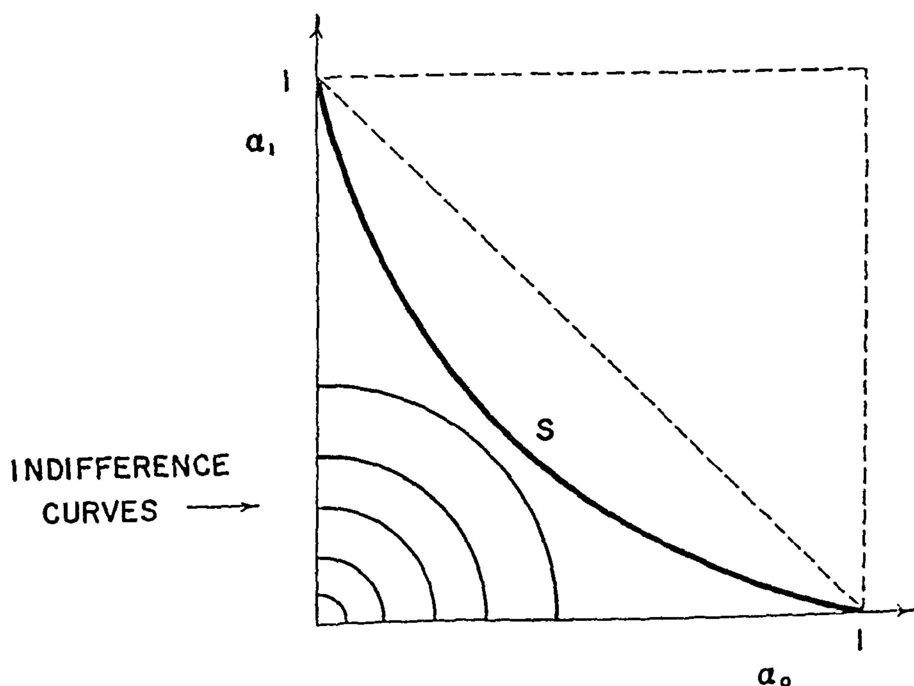


FIG. 1

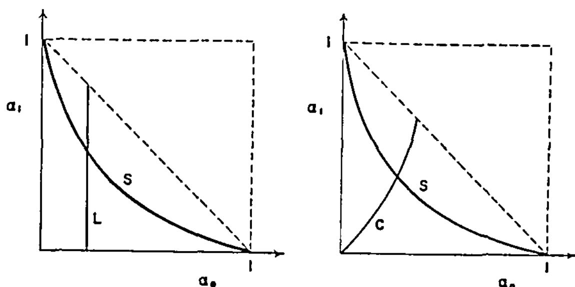


FIG. 2

intersection of S and L. This procedure is commonly justified on the grounds that the error of the *first kind is of a higher order of importance, and should* therefore be controlled at the prescribed level. However, if the curve S is sufficiently close to the α_0- and α_1-axis, as will always be the case if the sample size is sufficiently large, then α_1 is much smaller than α_0, which is inconsistent with the assumed relative importance of the two errors.

A more reasonable solution is obtained if one replaces the vertical line segment L by a curve $C: \alpha_1 = f(\alpha_0)$ where f is a continuous strictly increasing function with $f(0) = 0$. A particularly simple choice for f is a linear function

$$\alpha_1 = k\alpha_0 \tag{9}$$

Since $\alpha_0 \leqq 1 - \alpha_1$ for all admissible tests, one has $\alpha_0 \leqq 1/(k + 1)$ so that $1/(k + 1)$ *is an upper bound for* α_0. *As an example, consider* (9) *with* $k = 9$. If $\beta_1 = 1 - \alpha_1$ denotes the power of a test against the alternative θ_1, some typical pairs of values of (α_0, β_1) are

α_0	.1	.05	.04	.03	.02	.01	.005
β_1	.1	.55	.64	.73	.82	.91	.955

with .1 being an upper bound for α_0.

One would of course hope to avoid cases such as $\alpha_0 = .1$, $\beta_1 = .1$ or even $\alpha_0 = .05$, $\beta_1 = .55$. When no nuisance parameters are present, this can be achieved by taking a sample of sufficient size. In the composite case, on the other hand, *it can frequently not be achieved by samples of fixed size no matter how large*, but only by resorting to sequential experimentation.

To avoid misunderstandings, it should be emphasized that (9) is not being proposed as a logically convincing rule, nor as one fitting all occasions. Actually, it seems clear that no rule satisfying these requirements exists, except the Bayes

solution when sufficient knowledge concerning losses and *a priori* probabilities is available. In the absence of this knowledge it may be convenient to employ a simple rule of thumb. Such a rule is in fact being used in much of present practice: It consists in choosing α to be .05 or .01 depending on the seriousness attached to the committing of an error of the first kind. To this, (9) is suggested as an alternative which appears to be more reasonable in many cases.

It so happens that (9) is the minimax solution if the loss for rejecting $H:\theta \leqq \theta_0$ is a_0 when H is true, and the loss is a_1 for accepting H when $\theta \geqq \theta_1$, where the constant k of (9) is then given by $k = a_0/a_1$. However, this is not the basis for the present suggestion of (9), and the minimax property does not carry over to the application to be made in the next section to composite hypotheses.

4. Conditional tests. We return now to the composite case of the exponential family (1) with $r > 0$. The minimal complete class $\mathcal{C}$ is then more complex than in the preceding section, its members being characterized by the function $\alpha(\vartheta)$ instead of the single number α_0. Given any function $\alpha(\vartheta)$, which is the expectation of some critical function ϕ, there exists a unique member of $\mathcal{C}$ whose expectation function for $\theta = \theta_0$ is also $\alpha(\vartheta)$. This uniformly minimizes the risk (and maximizes the power) among all critical functions having this expectation.

If the alternatives of interest are as before those satisfying $\theta \geqq \theta_1$, let $\beta(\vartheta)$ denote the power function of a test against the alternative (θ_1, ϑ). The proposal made in the preceding section suggests selecting that member of $\mathcal{C}$ which satisfies

$$(10) \qquad 1 - \beta(\vartheta) = k\alpha(\vartheta) \qquad \text{for all } \vartheta.$$

However, this relationship depends on the particular parametrization chosen, and we shall not discuss it here. Instead an alternative approach will be proposed in which this difficulty does not arise.

Consider once more the case of the similar test with $\alpha(\vartheta) \equiv \alpha$. Since T is a complete sufficient statistic for ϑ when $\theta = \theta_0$, the functions C and γ of (2) are determined by the requirement that the conditional probability of rejection

$$\alpha^*(t) = P_{\theta_0}\{U > C(t) \mid t\} + \gamma(t)P_{\theta_0}\{U = C(t) \mid t\}$$

be equal to α for all t.[4] However, the conditional power $\beta^*(t) = P_{\theta_1}\{\text{rejecting } H \mid t\}$ of the test against the alternative $\theta = \theta_1$, typically depends on t. The question then arises: Suppose that $\beta^*(t)$ is quite small for the observed t, or quite high; is this value not more relevant to the case in hand than the average value $\beta(\vartheta)$?

Without entering into the difficulties raised by this question, there is an alternative and simpler justification for considering $\beta^*(t)$. The actual power β against the alternative $\theta = \theta_1$ generally depends on the nuisance parameter ϑ and is therefore unknown. It can however be estimated from the observations,

[4] This method of constructing exact tests was originated by Bartlett [1] and Neyman [9]. That in the present case it provides the totality of such tests has been noted by many authors. For a recent discussion and references see [7].

and $\beta^*(T)$ is the unbiased estimate with (uniformly) minimum variance. That it is unbiased is clear since $\beta(\vartheta) = E_{\theta_1,\vartheta}\beta^*(T)$. The minimum variance property is an immediate consequence of the completeness of the sufficient statistic T for (θ_1, ϑ) and of Theorem 5.1 of [7].

Analogous remarks apply in the more general case, in which the tests are not required to be exact. If the relevant frame of reference is obtained by considering t as fixed, the error probabilities of interest are the conditional probabilities $\alpha_0^*(t) = P_{\theta_0}$ (rejecting $H \mid t$) and $\alpha_1^*(t) = P_{\theta_1}$ (accepting $H \mid t$), and the quantities $C(t)$ and $\gamma(t)$ can therefore be determined from the relation

$$\alpha_1^*(t) = k\alpha_0^*(t). \tag{11}$$

The resulting test will of course not be similar. However, since $\alpha_0^*(t) \leq 1/(k+1)$ for all t, the quantity $1/(k+1)$ is an upper bound also for the average probability $\alpha_0(\vartheta)$ of an error of the first kind.

The above discussion applies only to problems in which the parameter of interest is one of the "natural" parameters of the exponential distribution (1). As was pointed out in [7], any parameter of the form $\theta + \sum a_i\vartheta_i$ is natural for a suitable definition of U, the T's and ϑ's. When the parameter of interest is not of this form, related methods may be applicable as is indicated by the following example.

If $X_1, \cdots, X_n$ are a sample from a normal distribution $N(\xi, \sigma^2)$, neither the parameter ξ nor ξ/σ are of this form. The problem of testing $\xi/\sigma \leq \delta_0$ against $\xi/\sigma \geq \delta_1$ can be reduced by invariance considerations to the statistic $\bar{X}/[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}$, the distribution of which depends on the single parameter $\delta = \xi/\sigma$. If $\alpha_i = P_{\delta_i}\{\bar{X} > C[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}\}$, the quantity C can be determined so that $\alpha_1 = k\alpha_0$. The problem of testing $\xi \leq \xi_0$ against $\xi \geq \xi_1$ appears to be more difficult; a possible approach may be that of [4], Section 3.

5. Examples. We shall now briefly indicate some examples in which the natural parameter θ is the relevant one so that the method of the preceding sections is applicable. Of these, Examples 1, 2, 3 have been treated by the same method (but from a different point of view) by Tocher [12], and Examples 2, 3 by Sverdrup [11].

EXAMPLE 1. Let X, Y be independent Poisson variables with $E(X) = \lambda$, $E(Y) = \mu$, and consider the problem of testing $\mu/\lambda \leq a_0$ against $\mu/\lambda \geq a_1$. The joint distribution of X, Y forms an exponential family with $T = X + Y$, $U = Y$, $\theta = \log(\mu/\lambda)$ and $\vartheta = \log \lambda$. The conditional distribution of Y given $X + Y = t$ is a binomial distribution corresponding to the success probability $p = \mu/(\lambda + \mu)$ and number of trials equal to t. In terms of p, the hypothesis and class of alternatives becomes $p \leq a_0/(a_0 + 1)$ and $p \geq a_1/(a_1 + 1)$ so that the test satisfying (2) and (11) can be determined from a table of the binomial distribution.

EXAMPLE 2. If X, Y are independent variables with binomial distributions $b(p_1, m)$ and $b(p_2, n)$, their joint distribution has the exponential form (1) with

solution when sufficient knowledge concerning losses and *a priori* probabilities is available. In the absence of this knowledge it may be convenient to employ a simple rule of thumb. Such a rule is in fact being used in much of present practice: It consists in choosing α to be .05 or .01 depending on the seriousness attached to the committing of an error of the first kind. To this, (9) is suggested as an alternative which appears to be more reasonable in many cases.

It so happens that (9) is the minimax solution if the loss for rejecting $H:\theta \leqq \theta_0$ is a_0 when H is true, and the loss is a_1 for accepting H when $\theta \geqq \theta_1$, where the constant k of (9) is then given by $k = a_0/a_1$. However, this is not the basis for the present suggestion of (9), and the minimax property does not carry over to the application to be made in the next section to composite hypotheses.

4. Conditional tests. We return now to the composite case of the exponential family (1) with $r > 0$. The minimal complete class $\mathcal{C}$ is then more complex than in the preceding section, its members being characterized by the function $\alpha(\vartheta)$ instead of the single number α_0. Given any function $\alpha(\vartheta)$, which is the expectation of some critical function ϕ, there exists a unique member of $\mathcal{C}$ whose expectation function for $\theta = \theta_0$ is also $\alpha(\vartheta)$. This uniformly minimizes the risk (and maximizes the power) among all critical functions having this expectation.

If the alternatives of interest are as before those satisfying $\theta \geqq \theta_1$, let $\beta(\vartheta)$ denote the power function of a test against the alternative (θ_1, ϑ). The proposal made in the preceding section suggests selecting that member of $\mathcal{C}$ which satisfies

$$1 - \beta(\vartheta) = k\alpha(\vartheta) \qquad \text{for all } \vartheta. \tag{10}$$

However, this relationship depends on the particular parametrization chosen, and we shall not discuss it here. Instead an alternative approach will be proposed in which this difficulty does not arise.

Consider once more the case of the similar test with $\alpha(\vartheta) \equiv \alpha$. Since T is a complete sufficient statistic for ϑ when $\theta = \theta_0$, the functions C and γ of (2) are determined by the requirement that the conditional probability of rejection

$$\alpha^*(t) = P_{\vartheta_0}\{U > C(t) \mid t\} + \gamma(t)P_{\vartheta_0}\{U = C(t) \mid t\}$$

be equal to α for all t.[4] However, the conditional power $\beta^*(t) = P_{\theta_1}\{\text{rejecting } H \mid t\}$ of the test against the alternative $\theta = \theta_1$, typically depends on t. The question then arises: Suppose that $\beta^*(t)$ is quite small for the observed t, or quite high; is this value not more relevant to the case in hand than the average value $\beta(\vartheta)$?

Without entering into the difficulties raised by this question, there is an alternative and simpler justification for considering $\beta^*(t)$. The actual power β against the alternative $\theta = \theta_1$ generally depends on the nuisance parameter ϑ and is therefore unknown. It can however be estimated from the observations,

[4] This method of constructing exact tests was originated by Bartlett [1] and Neyman [9]. That in the present case it provides the totality of such tests has been noted by many authors. For a recent discussion and references see [7].

and $\beta^*(T)$ is the unbiased estimate with (uniformly) minimum variance. That it is unbiased is clear since $\beta(\vartheta) = E_{\theta_1,\vartheta}\beta^*(T)$. The minimum variance property is an immediate consequence of the completeness of the sufficient statistic T for (θ_1, ϑ) and of Theorem 5.1 of [7].

Analogous remarks apply in the more general case, in which the tests are not required to be exact. If the relevant frame of reference is obtained by considering t as fixed, the error probabilities of interest are the conditional probabilities $\alpha_0^*(t) = P_{\theta_0}$ (rejecting $H \mid t$) and $\alpha_1^*(t) = P_{\theta_1}$ (accepting $H \mid t$), and the quantities $C(t)$ and $\gamma(t)$ can therefore be determined from the relation

$$\alpha_1^*(t) = k\alpha_0^*(t). \tag{11}$$

The resulting test will of course not be similar. However, since $\alpha_0^*(t) \leqq 1/(k+1)$ for all t, the quantity $1/(k+1)$ is an upper bound also for the average probability $\alpha_0(\vartheta)$ of an error of the first kind.

The above discussion applies only to problems in which the parameter of interest is one of the "natural" parameters of the exponential distribution (1). As was pointed out in [7], any parameter of the form $\theta + \sum a_i\vartheta_i$ is natural for a suitable definition of U, the T's and ϑ's. When the parameter of interest is not of this form, related methods may be applicable as is indicated by the following example.

If $X_1, \cdots, X_n$ are a sample from a normal distribution $N(\xi, \sigma^2)$, neither the parameter ξ nor ξ/σ are of this form. The problem of testing $\xi/\sigma \leqq \delta_0$ against $\xi/\sigma \geqq \delta_1$ can be reduced by invariance considerations to the statistic $\bar{X}/[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}$, the distribution of which depends on the single parameter $\delta = \xi/\sigma$. If $\alpha_i = P_{\delta_i}\{\bar{X} > C[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}\}$ the quantity C can be determined so that $\alpha_1 = k\alpha_0$. The problem of testing $\xi \leqq \xi_0$ against $\xi \geqq \xi_1$ appears to be more difficult; a possible approach may be that of [1], Section 3.

5. Examples. We shall now briefly indicate some examples in which the natural parameter θ is the relevant one so that the method of the preceding sections is applicable. Of these, Examples 1, 2, 3 have been treated by the same method (but from a different point of view) by Tocher [12], and Examples 2, 3 by Sverdrup [11].

Example 1. Let X, Y be independent Poisson variables with $E(X) = \lambda$, $E(Y) = \mu$, and consider the problem of testing $\mu/\lambda \leqq a_0$ against $\mu/\lambda \geqq a_1$. The joint distribution of X, Y forms an exponential family with $T = X + Y$, $U = Y$, $\theta = \log(\mu/\lambda)$ and $\vartheta = \log\lambda$. The conditional distribution of Y given $X + Y = t$ is a binomial distribution corresponding to the success probability $p = \mu/(\lambda + \mu)$ and number of trials equal to t. In terms of p, the hypothesis and class of alternatives becomes $p \leqq a_0/(a_0 + 1)$ and $p \geqq a_1/(a_1 + 1)$ so that the test satisfying (2) and (11) can be determined from a table of the binomial distribution.

Example 2. If X, Y are independent variables with binomial distributions $b(p_1, m)$ and $b(p_2, n)$, their joint distribution has the exponential form (1) with

solution when sufficient knowledge concerning losses and *a priori* probabilities is available. In the absence of this knowledge it may be convenient to employ a simple rule of thumb. Such a rule is in fact being used in much of present practice: It consists in choosing α to be .05 or .01 depending on the seriousness attached to the committing of an error of the first kind. To this, (9) is suggested as an alternative which appears to be more reasonable in many cases.

It so happens that (9) is the minimax solution if the loss for rejecting $H:\theta \leqq \theta_0$ is a_0 when H is true, and the loss is a_1 for accepting H when $\theta \geqq \theta_1$, where the constant k of (9) is then given by $k = a_0/a_1$. However, this is not the basis for the present suggestion of (9), and the minimax property does not carry over to the application to be made in the next section to composite hypotheses.

4. Conditional tests. We return now to the composite case of the exponential family (1) with $r > 0$. The minimal complete class $\mathcal{C}$ is then more complex than in the preceding section, its members being characterized by the function $\alpha(\vartheta)$ instead of the single number α_0. Given any function $\alpha(\vartheta)$, which is the expectation of some critical function ϕ, there exists a unique member of $\mathcal{C}$ whose expectation function for $\theta = \theta_0$ is also $\alpha(\vartheta)$. This uniformly minimizes the risk (and maximizes the power) among all critical functions having this expectation.

If the alternatives of interest are as before those satisfying $\theta \geqq \theta_1$, let $\beta(\vartheta)$ denote the power function of a test against the alternative (θ_1, ϑ). The proposal made in the preceding section suggests selecting that member of $\mathcal{C}$ which satisfies

$$1 - \beta(\vartheta) = k\alpha(\vartheta) \qquad \text{for all } \vartheta. \tag{10}$$

However, this relationship depends on the particular parametrization chosen, and we shall not discuss it here. Instead an alternative approach will be proposed in which this difficulty does not arise.

Consider once more the case of the similar test with $\alpha(\vartheta) \equiv \alpha$. Since T is a complete sufficient statistic for ϑ when $\theta = \theta_0$, the functions C and γ of (2) are determined by the requirement that the conditional probability of rejection

$$\alpha^*(t) = P_{\theta_0}\{U > C(t) \mid t\} + \gamma(t)P_{\theta_0}\{U = C(t) \mid t\}$$

be equal to α for all t.[4] However, the conditional power $\beta^*(t) = P_{\theta_1}\{\text{rejecting } H \mid t\}$ of the test against the alternative $\theta = \theta_1$, typically depends on t. The question then arises: Suppose that $\beta^*(t)$ is quite small for the observed t, or quite high; is this value not more relevant to the case in hand than the average value $\beta(\vartheta)$?

Without entering into the difficulties raised by this question, there is an alternative and simpler justification for considering $\beta^*(t)$. The actual power β against the alternative $\theta = \theta_1$ generally depends on the nuisance parameter ϑ and is therefore unknown. It can however be estimated from the observations,

[4] This method of constructing exact tests was originated by Bartlett [1] and Neyman [9]. That in the present case it provides the totality of such tests has been noted by many authors. For a recent discussion and references see [7].

and $\beta^*(T)$ is the unbiased estimate with (uniformly) minimum variance. That it is unbiased is clear since $\beta(\vartheta) = E_{\theta_1,\vartheta}\beta^*(T)$. The minimum variance property is an immediate consequence of the completeness of the sufficient statistic T for (θ_1, ϑ) and of Theorem 5.1 of [7].

Analogous remarks apply in the more general case, in which the tests are not required to be exact. If the relevant frame of reference is obtained by considering t as fixed, the error probabilities of interest are the conditional probabilities $\alpha_0^*(t) = P_{\theta_0}$ (rejecting $H \mid t$) and $\alpha_1^*(t) = P_{\theta_1}$ (accepting $H \mid t$), and the quantities $C(t)$ and $\gamma(t)$ can therefore be determined from the relation

$$\alpha_1^*(t) = k\alpha_0^*(t) \tag{11}$$

The resulting test will of course not be similar. However, since $\alpha_0^*(t) \leqq 1/(k+1)$ for all t, the quantity $1/(k+1)$ is an upper bound also for the average probability $\alpha_0(\vartheta)$ of an error of the first kind.

The above discussion applies only to problems in which the parameter of interest is one of the "natural" parameters of the exponential distribution (1). As was pointed out in [7], any parameter of the form $\theta + \sum a_i\vartheta_i$ is natural for a suitable definition of U, the T's and ϑ's. When the parameter of interest is not of this form, related methods may be applicable as is indicated by the following example.

If $X_1, \cdots, X_n$ are a sample from a normal distribution $N(\xi, \sigma^2)$, neither the parameter ξ nor ξ/σ are of this form. The problem of testing $\xi/\sigma \leqq \delta_0$ against $\xi/\sigma \geqq \delta_1$ can be reduced by invariance considerations to the statistic $\bar{X}/[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}$, the distribution of which depends on the single parameter $\delta = \xi/\sigma$. If $\alpha_i = P_{\delta_i}\{\bar{X} > C[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}\}$, the quantity C can be determined so that $\alpha_1 = k\alpha_0$. The problem of testing $\xi \leqq \xi_0$ against $\xi \geqq \xi_1$ appears to be more difficult; a possible approach may be that of [4], Section 3.

6. Examples. We shall now briefly indicate some examples in which the natural parameter θ is the relevant one so that the method of the preceding sections is applicable. Of these, Examples 1, 2, 3 have been treated by the same method (but from a different point of view) by Tocher [12], and Examples 2, 3 by Sverdrup [11].

EXAMPLE 1. Let X, Y be independent Poisson variables with $E(X) = \lambda$, $E(Y) = \mu$, and consider the problem of testing $\mu/\lambda \leqq a_0$ against $\mu/\lambda \geqq a_1$. The joint distribution of X, Y forms an exponential family with $T = X + Y$, $U = Y$, $\theta = \log(\mu/\lambda)$ and $\vartheta = \log \lambda$. The conditional distribution of Y given $X + Y = t$ is a binomial distribution corresponding to the success probability $p = \mu/(\lambda + \mu)$ and number of trials equal to t. In terms of p, the hypothesis and class of alternatives becomes $p \leqq a_0/(a_0 + 1)$ and $p \geqq a_1/(a_1 + 1)$, so that the test satisfying (2) and (11) can be determined from a table of the binomial distribution.

EXAMPLE 2. If X, Y are independent variables with binomial distributions $b(p_1, m)$ and $b(p_2, n)$, their joint distribution has the exponential form (1) with

$T = X + Y$, $U = Y$, $\theta = \log (p_2/q_2 \div p_1/q_1)$ and $\vartheta = \log (p_1/q_1)$. The method is therefore applicable to the problem of testing $p_2/q_2 \leqq a_0(p_1/q_1)$, and in particular $p_2 \leqq p_1$ by letting $a_0 = 1$, against the alternatives $p_2/q_2 \geqq a_1(p_1/q_1)$. Putting $\rho = (p_2/q_2) \div (p_1/q_1)$, the conditional distribution of Y given t is

$$(12) \qquad P_\rho\{Y = y \mid X + Y = t\} = C_t(\rho) \binom{m}{t - y} \binom{n}{y} \rho^y, \qquad y = 0, 1, \cdots, t,$$

which for $\rho = 1$ reduces to the hypergeometric distribution.

Example 3. In a 2×2 table representing the results of classifying a sample of size s according to two characteristics A and B, the joint distribution of the numbers X, Y, Y' in the

	A	$\bar{A}$	
B	X	X'	M
$\bar{B}$	Y	Y'	N
	T	T'	S

categories AB, $A\bar{B}$ and $\bar{A}\bar{B}$ constitute an exponential family with $U = Y$, $T_1 = X + Y$, $T_2 = Y + Y'$ and $\theta = \log (p_{A\bar{B}}p_{\bar{A}B}/p_{AB}p_{\bar{A}\bar{B}})$. Putting $\Delta = (p_{A\bar{B}}p_{\bar{A}B}/p_{AB}p_{\bar{A}\bar{B}})$ one finds

$$p_{AB} = p_A p_B + \frac{1 - \Delta}{\Delta} p_{\bar{A}B} p_{A\bar{B}}; \qquad p_{\bar{A}\bar{B}} = p_{\bar{A}} p_{\bar{B}} + \frac{1 - \Delta}{\Delta} p_{\bar{A}B} p_{A\bar{B}}$$

$$p_{A\bar{B}} = p_A p_{\bar{B}} - \frac{1 - \Delta}{\Delta} p_{\bar{A}B} p_{A\bar{B}}; \qquad p_{\bar{A}B} = p_{\bar{A}} p_B - \frac{1 - \Delta}{\Delta} p_{\bar{A}B} p_{A\bar{B}}$$

where p_{AB} denotes the probability of having the characteristics A and B, $p_A = p_{AB} + p_{A\bar{B}}$ the probability of having the characteristic A, etc. The quantity Δ is therefore a measure of the degree of dependence,[5] $\Delta = 1$ corresponding to independence, $\Delta < 1$ to negative and $\Delta > 1$ to positive dependence. The method of the preceding section is applicable to testing $\Delta \leqq 1$ or more generally $\Delta \leqq \Delta_0$ against the alternatives $\Delta \geqq \Delta_1$. The conditional distribution of Y given $X + Y = t$, $Y + Y' = n$ is given by (12) with Δ in place of ρ.

Example 4. Consider a number of paired comparisons (U_k, V_k) where only the sign of the differences $W_k = V_k - U_k$ are observed for each pair $k = 1, \cdots, n$. If the probability of a positive, negative and zero observation are p_+, p_- and p_0 in each case and if the comparisons are independent, the joint distribution of the numbers X, Y and Z of positive, negative and zero cases is the multinomial distribution

$$\frac{n!}{x!y!z!} p_+^x p_-^y p_0^z.$$

[5] Δ is equivalent to Yule's measure of association, which is $Q = (1 - \Delta)/(1 + \Delta)$. For a discussion of this and related measures, see [2].

This is an exponential family with $U = Y$, $T = Z$, $\theta = \log(p_+/p_-)$ and $\vartheta = \log(p_0/p_-)$. The test of $p_+ \leqq p_-$ (or $p_+ \leqq a_0 p_-$) against $p_+ \geqq a_1 p_-$ is therefore performed conditionally given $Z = t$. Since the conditional distribution of Y given $Z = t$ is the binomial distribution $b(p_+/(p_+ + p_-), n - t)$, the constants $C(t)$ *and* $\gamma(t)$ *for which the test satisfies* (2) and (11) can be obtained from the binomial tables.[6]

EXAMPLE 5. Let $Y_1, \cdots, Y_N$ be independently distributed according to the binomial distributions $b(p_i, n_i)$ $i = 1, \cdots, N$ where

$$p_i = 1/[1 + e^{-(\alpha+\beta x_i)}]$$

This is the model frequently assumed in bioassay, where x_i denotes the dose or some function of the dose such as its logarithm, of a drug given to n_i experimental subjects and where Y_i is the number among these subjects which respond to the drug at level x_i. Here the x_i are known, and α and β are unknown parameters. The joint distribution of the Y's is

$$e^{\alpha \Sigma y_i + \beta \Sigma x_i y_i} \prod_{i=1}^{N} \binom{n_i}{y_i} \left[\frac{e^{-(\alpha+\beta x_i)}}{1 + e^{-(\alpha+\beta x_i)}} \right]^{n_i}, \tag{13}$$

which is an exponential family with the parameters α, β and sufficient statistics $\sum Y_i$, $\sum x_i Y_i$. The method is therefore applicable to testing $\alpha \leqq \alpha_0$ against $\alpha \geqq \alpha_1$ or $\beta \leqq \beta_0$ against $\beta \geqq \beta_1$. It is interesting to note that for the particular case $x_i = ic$ and $H: \beta \leqq 0$, the conditional test given $Y = t$ is a form of the Wilcoxon test in a setting similar to that discussed by Haldane and Smith [3]

As a last example we mention without going into details the comparison of two distributions of type (13). If the parameters in these are α, β and α', β' the differences $\alpha' - \alpha$ and $\beta' - \beta$ are natural parameters of the resulting exponential families, and can therefore be tested by the method discussed here

REFERENCES

[1] M. S. BARTLETT, "Properties of sufficiency and statistical tests," *Proc Roy Soc. London*, Vol. 160 (1937), pp. 268-281.

[2] LEO A. GOODMAN AND WILLIAM H. KRUSKAL, "Measures of association for mass classifications," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 732-764

[3] J. B. S. HALDANE AND CEDRIC A. B. SMITH, "A simple exact test for birth order effect," *Ann. Eugenics*, Vol. 14 (1948), pp. 117-124.

[4] J. L. HODGES, JR. AND E. L. LEHMANN, "Testing the approximate validity of statistical hypotheses," *J. Roy. Stat. Soc., Ser. B.*, Vol. 16 (1954), pp. 261-268.

[5] H JEFFREYS, "Theory of probability (2d ed.), Oxford: Clarendon Press, 1948

[6] SAMUEL KARLIN AND HERMAN RUBIN, "The theory of decision procedures for distribu

[6] This problem has been considered previously in [10] The statement made there that the test satisfying (2) and $\alpha^*(t) = \alpha$ is uniformly most powerful is too strong. The test is however uniformly most powerful among all similar (and hence all unbiased) tests An analogous remark applies to Example 2, which is among those considered in [12] As was proved there, the test which is conditionally unbiased at fixed level α for each t, is uniformly unbiased; it is however not uniformly most powerful without this restriction as is claimed in [12] for all the cases treated there.

tions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272-299.
[7] E. L. LEHMANN AND HENRY SCHEFFÉ, "Completeness, similar regions and unbiased estimation," *Sankhya*, Vol. 10 (1950), pp. 305-340 and Vol. 15 (1955), pp. 219-236.
[8] D. V. LINDLEY, "Statistical inference," *J. Roy. Stat. Soc., Ser. B.*, Vol. 15 (1953), pp. 30-76.
[9] J. NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability," *Philos. Trans. Roy. Soc. London*, Vol. 236 (1937), pp. 333-380.
[10] JOSEPH PUTTER, "The treatment of ties in some nonparametric tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 368-386.
[11] ERLING SVERDRUP, "Similarity, unbiasedness, minimaxibility, and admissibility of statistical test procedures," *Skand. Aktuarietids.*, Vol. 36 (1953), pp. 64-86.
[12] K. D. TOCHER, "Extension of Neyman-Pearson theory of tests to discontinuous variates," *Biometrika*, Vol. 37 (1950), pp. 130-144.
[13] DONALD R. TRUAX, "Multidecision problems for the multivariate exponential family," Stanford Technical Report No. 32 (1955) (to be published).

STEP-DOWN PROCEDURE IN MULTIVARIATE ANALYSIS[1]

By J. Roy

University of North Carolina

1. Introduction and summary. Test criteria for (i) multivariate analysis of variance, (ii) comparison of variance-covariance matrices, and (iii) multiple independence of groups of variates when the parent population is multivariate normal are usually derived either from the likelihood-ratio principle [6] or from the "union-intersection" principle [2]. An alternative procedure, called the "step-down" procedure, has been recently used by Roy and Bargmann [5] in devising a test for problem (iii). In this paper the step-down procedure is applied to problems (i) and (ii) in deriving new tests of significance and simultaneous confidence-bounds on a number of "deviation-parameters."

The essential point of the step-down procedure in multivariate analysis is that the variates are supposed to be arranged in descending order of importance. The hypothesis concerning the multivariate distribution is then decomposed into a number of hypotheses—the first hypothesis concerning the marginal univariate distribution of the first variate, the second hypothesis concerning the conditional univariate distribution of the second variate given the first variate, the third hypothesis concerning the conditional univariate distribution of the third variate given the first two variates, and so on. For each of these component hypotheses concerning univariate distributions, well known test procedures with good properties are usually available, and these are made use of in testing the compound hypothesis on the multivariate distribution. The compound hypothesis is accepted if and only if each of the univariate hypotheses are accepted. It so turns out that the component univariate tests are independent, if the compound hypothesis is true. It is therefore possible to determine the level of significance of the compound test in terms of the levels of significance of the component univariate tests and to derive simultaneous confidence-bounds on certain meaningful parametric functions on the lines of [3] and [4].

The step-down procedure obviously is not invariant under a permutation of the variates and should be used only when the variates can be arranged on a priori grounds. Some advantages of the step-down procedure are (i) the procedure uses widely known statistics like the variance-ratio, (ii) the test is carried out in successive stages and if significance is established at a certain stage, one can stop at that stage and no further computations are needed, and (iii) it leads to simultaneous confidence-bounds on certain meaningful parametric functions.

1.1 *Notations.* The operator $\mathcal{E}$ applied to a matrix of random variables is used

Received January 17, 1958; revised May 5, 1958.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command. Reproduction in whole or in part is permitted for any purpose of the United States Government.

to generate the matrix of expected values of the corresponding random variables. The form of a matrix is denoted by a subscript; thus $A_{n \times m}$ indicates that the matrix A has n rows and m columns. The maximum latent root of a square matrix B is denoted by $\lambda_{\max}(B)$. Given a vector $a = (a_1, a_2, \cdots, a_t)'$ and a subset T of the natural numbers 1, 2, $\cdots$, t, say $T = (j_1, j_2, \cdots, j_u)$ where $j_1 < j_2 < \cdots j_u$, the notation $T[a]$ will be used to denote the positive quantity:

$$T[a] = + \{a_{j_1}^2 + a_{j_2}^2 + \cdots + a_{j_t}^2\}^{1/2}.$$

$T[a]$ will be called the T-norm of a. Similarly, given a matrix $B_{t \times t}$, we shall write $B_{(T)}$ for the $u \times u$ submatrix formed by taking the j_1th, j_2th, $\cdots$, j_uth rows and columns of B. We shall call $B_{(T)}$ the T-submatrix of B.

2. Step-down procedure in multivariate analysis of variance.

2.1 *General linear hypothesis in univariate analysis.* Let the elements of $y_{n \times 1}$ be one-dimensional random variables distributed independently and normally with the same variance σ^2 and expectations given by

$$\mathcal{E}y = A\theta + X\beta \tag{1}$$

where elements of $\theta_{m \times 1}$ and $\beta_{q \times 1}$ are unknown parameters; $A_{n \times m}$ and $X_{n \times q}$ are matrices of known constants with rank $(A) = r$ and rank $(A:X) = r + q$, with $n > (r + q)$.

A set of t linearly independent linear functions $\phi_{t \times 1} = B_{t \times m}\theta$, where B is a given matrix of rank t, is said to be estimable if for each element of ϕ there exists an unbiassed estimate linear in y, for all values of θ and β. If ϕ is estimable, there exists an estimator $\hat{\phi}_{t \times 1}$ of ϕ, the elements of which are linear in y and minimum variance unbiassed estimators of the corresponding elements in ϕ. Denote the variance-covariance matrix of $\hat{\phi}$ by $C \cdot \sigma^2$, where $C_{t \times t}$ is a positive-definite matrix. Let $s^2/(n - q - r)$ denote the usual error mean square with $(n - q - r)$ degrees of freedom giving an unbiassed estimator of σ^2. Then it is well known that the statistics $u = (\hat{\phi} - \phi)'C^{-1}(\hat{\phi} - \phi)/\sigma^2$ and $v = s^2/\sigma^2$ are distributed independently as chi-squares with t and $(n - q - r)$ degrees of freedom respectively, so that

$$F \equiv \frac{(\hat{\phi} - \phi)'C^{-1}(\hat{\phi} - \phi)/t}{s^2/(n - q - r)} \tag{2}$$

is distributed as a variance-ratio with t and $(n - q - r)$ degrees of freedom.

Let α be a preassigned constant, $0 < \alpha < 1$, and f the upper 100α per cent point of the variance-ratio distribution with t and $(n - q - r)$ degrees of freedom. Setting $\mathfrak{L}^2 = tf/(n - q - r)$ we then have

$$(\hat{\phi} - \phi)'C^{-1}(\hat{\phi} - \phi) \leqq \ell^2 s^2 \tag{3}$$

with probability $(1 - \alpha)$.

Now, the left-hand side of (3) is a positive definite quadratic form in $(\hat{\phi} - \phi)$ and consequently, we have

$$(\hat{\phi} - \phi)'C^{-1}(\hat{\phi} - \phi) \geqq (\hat{\phi} - \phi)'(\hat{\phi} - \phi)/\lambda_{\max}(C). \tag{4}$$

We thus have

(5) $$(\hat{\phi} - \phi)'(\hat{\phi} - \phi) \leqq t^2 s^2 \lambda_{max}(C)$$

with probability not less than $(1 - \alpha)$.

Now, let T be any subset of the natural numbers $1, 2, \cdots, t$ and consider the T-norms $T[\phi]$ of ϕ and $T[\hat{\phi}]$ of $\hat{\phi}$. Then (3) implies that

(6) $$T[\hat{\phi}] - ts\,\lambda_{max}^{1/2}(C_{(T)}) \leqq T[\phi] \leqq T[\hat{\phi}] + ts\,\lambda_{max}^{1/2}(C_{(T)})$$

for all subsets T of $(1, 2, \cdots, t)$, where C_T is the T-submatrix of C. The statement (6) thus provides simultaneous confidence-bounds on the parameters $T[\phi]$ for all T with probability not less than $(1 - \alpha)$. We note that there are in all $(2^t - 1)$ parameters of the type $T[\phi]$ and these in a sense measure the deviations from the hypothesis $\mathcal{H}_0$ that $\phi = 0$. The analysis of variance test for $\mathcal{H}_0$ at level of significance α, of course, is given by the rule

(7) $$\text{accept } \mathcal{H}_0 \text{ if } \frac{\hat{\phi}'C^{-1}\hat{\phi}/t}{s^2/(n - q - r)} \leqq f;$$
$$\text{otherwise reject } \mathcal{H}_0$$

However, simultaneous confidence-bounds of the type (6) are more interesting than the test (7) itself, because the direction of departure from the null hypothesis is indicated.

2.2 *Customary tests in multivariate analysis of variance.* We have a matrix $Y_{n \times p}$ of random variables, such that the rows are distributed independently, each row having a p-variate normal distribution with the same variance-covariance matrix $\Sigma_{p \times p}$ which is positive-definite. The expected values are given by

(8) $$\mathcal{E}Y = A\Theta,$$

where $A_{n \times m}$ is a matrix of known constants of rank r, $r \leqq (n - p)$, and $\Theta_{m \times p}$ is a matrix of unknown parameters. As before, a set of linear parametric functions $\Phi_{t \times p} = B_{t \times m}\Theta$ is said to be estimable if, for all Θ, there exist unbiased estimates of Φ linear in Y. If Φ is estimable, customary tests for the hypothesis

$$\mathcal{H}_0: \Phi = 0$$

are based on two $p \times p$ matrices of random variables

(9) $$S_e = Y'EY \quad \text{and} \quad S_h = Y'HY,$$

called respectively the sum of products matrix due to error and the sum of products matrix due to hypothesis. Here E and H are $n \times n$ symmetric idempotent matrices with non-stochastic elements, E of rank $(n - r)$ and H of rank t, E being a function of A, and H of both A and B. The likelihood-ratio test [6] is

(10) $$\text{accept } \mathcal{H}_0 \text{ if } L = \frac{|S_e|}{|S_e + S_h|} > c,$$
$$\text{otherwise reject } \mathcal{H}_0,$$

where c is a preassigned constant depending on the level of significance. The test based on the largest latent root [3] is

$$\text{accept } \mathcal{H}_0 \text{ if } \lambda_{\max}(S_h S_e^{-1}) < d, \qquad \text{otherwise reject } \mathcal{H}_0, \tag{11}$$

where d is a constant depending on the level of significance. Simultaneous confidence-bounds on certain meaningful parametric functions have been derived by the largest (or the largest-smallest roots) procedure, [3] [4], whereas no such bounds are available as of now from the likelihood-ratio procedure.

2.3 *The step-down procedure.* We shall denote the ith columns of the matrices Y and Θ in section 2.2 by y_i and θ_i respectively and write $Y_i = [y_1\, y_2\, \cdots\, y_i]$ and $\Theta_i = [\theta_1\, \theta_2\, \cdots\, \theta_i]$. Further, we shall denote the top left-hand $i \times i$ submatrix of $\Sigma \equiv ((\sigma_{ij}))$ by Σ_i.

Then, under the condition that Y_i is fixed, the n elements of the vector y_{i+1} are distributed independently and normally each with the same variance σ_{i+1}^2 and expectations given by

$$\mathcal{E}y_{i+1} = A\eta_{i+1} + Y_i\beta_i, \tag{12}$$

where β_i is a vector of the form $i \times 1$ given by

$$\beta_i = \Sigma_i^{-1}\begin{bmatrix}\sigma_{1,i+1}\\ \sigma_{2,i+1}\\ \cdots\\ \sigma_{i,i+1}\end{bmatrix}, \qquad \beta_0 = 0, \tag{13}$$

amd η_{i+1} is a vector of the form $m \times 1$ given by

$$\eta_{i+1} = \theta_{i+1} - \Theta_i\beta_i \tag{14}$$

and

$$\sigma_{i+1}^2 = \frac{|\Sigma_{i+1}|}{|\Sigma_i|}, \tag{15}$$

with the understanding that $|\Sigma_0| = 1$ so that $\sigma_1^2 = \sigma_{11}$, $i = 0, 1, 2, \cdots, (p-1)$. The elements of the vectors β_i, η_{i+1} may then be regarded as unknown parameters. We shall call β_i the ith order step-down regression coefficient and σ_{i+1}^2 the ith order step-down residual variance.

Let us now consider linear functions

$$\phi_i = B\eta_i \qquad (i = 1, 2, \cdots, p). \tag{16}$$

If Y_i is fixed, (12) is of the same form as (1). Let us now, with an easily understood notation similar to that used in Section 2.1, construct the statistics

$$F_i \equiv \frac{(\hat{\phi}_i - \phi_i)' C_i^{-1} (\hat{\phi}_i - \phi_i)'/t}{s_i^2/(n - r - i + 1)} \qquad (i = 1, 2, \cdots, p). \tag{17}$$

Obviously, when Y_{i-1} is fixed, the statistic F_i is distributed as a variance ratio with t and $(n - r - i + 1)$ degrees of freedom ($i = 2, 3, \cdots, p$). Finally, we note that in its functional form F_i involves only Y_i ($i = 1, 2, \cdots, p$) and that the conditional distribution of F_i, given Y_{i-1} does not involve Y_{i-1} ($i = 2, 3, \cdots, p$) and hence $F_{i-1}, \cdots, F_1$. Also, F_1 is marginally distributed as a variance-ratio with t and $(n - r)$ degrees of freedom Therefore the statistics $F_1, F_2, \cdots, F_p$ are independent. This can be verified in a straight-forward manner by using the transformation to rectangular coordinates as in [5] or any other set of step-down variates, or even otherwise.

For a preassigned constant α_i, $0 < \alpha_i < 1$, let f_i denote the upper $100\alpha_i$ per cent point of the variance-ratio distribution with t and $(n - r - i + 1)$ degrees of freedom. Then the probability P that simultaneously

$$F_i \leqq f_i, \qquad i = 1, 2, \cdots, p, \tag{18}$$

is given by

$$P = \prod_{i=1}^{p} (1 - \alpha_i) \tag{19}$$

Therefore, for any subset T of the natural numbers $1, 2, \cdots, t$ writing as in (6), $T[\phi_i]$ and $T[\dot{\phi}_i]$ for the T-norms of ϕ_i and $\dot{\phi}_i$ respectively, and setting

$$\ell_i^2 = t f_i / (n - r - i + 1) \tag{20}$$

and writing $C_{i(T)}$ for the T-submatrix of C_i, we have the simultaneous confidence bounds

$$T[\dot{\phi}_i] - \ell_i s_i \lambda_{\max}^{1/2}(C_{i(T)}) \leqq T[\phi_i] \leqq T[\dot{\phi}_i] + \ell_i s_i \lambda_{\max}^{1/2}(C_{i(T)}) \tag{21}$$

for all subsets T of $(1, 2, \cdots, t)$ and $i = 1, 2, \cdots, p$ with probability greater than P.

To derive a test of the hypothesis $\mathcal{H}_0$ that $\Phi = 0$, we note that $\mathcal{H}_0$ is true if and only if the hypothesis $\mathcal{H}_i$ that $\phi_i = 0$ holds for all $i = 1, 2, \cdots, p$ Using the result (17), we set up the following procedure for testing $\mathcal{H}_0$:

$$\text{(22)} \qquad \begin{aligned} &\text{accept } \mathcal{H}_0 \text{ if } u_i \equiv \frac{\dot{\phi}_i' C_i^{-1} \dot{\phi}_i / t}{s_{i\cdot}^2 / (n - r - i + 1)} \leqq f_i \qquad \text{for all } i = 1, 2, \cdots, p, \\ &\text{otherwise reject } \mathcal{H}_0. \end{aligned}$$

Obviously, the level of significance for this test is $1 - P$ where P is given by (19). The arbitrariness in determining the f_i's when the level of significance is preassigned may be removed by stipulating that $\alpha_1 = \alpha_2 = \cdots = \alpha_p$. From the fact that the variance-ratio test (7) is uniformly unbiased, it can be seen after a little consideration, that the test procedure (22) is also uniformly unbiased.

To carry out the test one should first compute u_1. If $u_1 > f_1$, $\mathcal{H}_0$ is rejected and no further computations are needed. *If $u_1 \leqq f_1$, the next step is to compute* u_2. If $u_2 > f_2$, $\mathcal{H}_0$ is rejected and no further computations are needed. If $u_2 \leqq f_2$,

one proceeds to compute u_3 and so on. This way one need compute u_i if and only if $u_j \leqq f_j$ for $j = 1, 2, \cdots, i-1$. Much computational labor is saved thereby.

It is well known that the likelihood-ratio statistic L given by (10) can be expressed as

$$L = \prod_{i=1}^{p} \frac{(n - r - i + 1)}{t + (n - r - i + 1)u_i} \tag{23}$$

and this has been utilized [1] to obtain the moments of L when $\mathcal{H}_0$ is true. However, the step-down procedure based on the individual u_i's rather than on a single function L, is advantageous from the point of view of (i) setting up simultaneous confidence bounds and (ii) saving computational labor, specially in the situation indicated in the introduction.

3. Step-down procedure for variance-covariance matrices. Let $S_{p \times p} \equiv ((s_{ij}))$ be a symmetric matrix of random variables, distributed in Wishart's form with n degrees of freedom, $n > p$, so that S/n provides an unbiassed estimate for the variance-covariance matrix Σ of a p-variate normal population. In the same way as in Section 2.3, we shall write S_i for the $i \times i$ top left-hand submatrix of S and let

$$b_i = S_i^{-1} \begin{bmatrix} s_{1,i+1} \\ s_{2,i+1} \\ \cdots \\ s_{i,i+1} \end{bmatrix}, \qquad b_0 = 0, \tag{24}$$

$$s_{i+1}^2 = \frac{|S_{i+1}|}{|S_i|}, \qquad s_1^2 = s_{11}, \tag{25}$$

for $i = 1, 2, \cdots, p-1$. Let β_{i-1} and σ_i^2 be defined by (13) and (15) for $i = 1, 2, \cdots, p$. Then it is well known that when S_i is fixed, the distribution of b_i is independent of the distribution of s_{i+1}^2; the distribution of b_i is i-variate normal with expectation β_i and variance-covariance matrix $\sigma_{i+1}^2 S_i^{-1}$, and s_{i+1}^2/σ_{i+1}^2 has the chi-square distribution with $(n-i)$ degrees of freedom, $i = 1, 2, \cdots, (p-1)$. Finally s_1^2/σ_1^2 has the chi-square distribution with n degrees of freedom.

When more than one variance-covariance matrix is involved, we shall distinguish them by a superscript under parentheses. Thus with a number of population variance-covariance matrices $\Sigma^{(j)}$ and the corresponding Wishart matrices $S^{(j)}$, the quantities $\beta_i^{(j)}$, $\sigma_i^{(j)}$, $b_i^{(j)}$, $s_i^{(j)}$, etc., will be defined in the same way as in (13), (15), (24), and (25) for $j = 1, 2, \cdots$, etc.

3.1 *One variance-covariance matrix.* On the basis of a matrix S distributed in Wishart's form with n degrees of freedom, with S/n providing an unbiassed estimate for Σ, it is possible to set up simultaneous confidence-bounds on parameters which are functions of the elements of Σ by the step-down procedure as follows.

When S_i is fixed, the statistics $u = (b_i - \beta_i)'S_i(b_i - \beta_i)/\sigma_{i+1}^2$ and $v =$

s_{i+1}^2/σ_{i+1}^2 are distributed independently as chi-squares, u with i degrees of freedom and v with $n - i$ degrees of freedom. Therefore, given pre-assigned positive constants a_i, c_{i+1}, and d_{i+1}, where $c_{i+1} < d_{i+1}$, the probability P_{i+1} that

(26)
$$(b_i - \beta_i)' S_i (b_i - \beta_i)/s_{i+1}^2 \leqq a_i^2, \qquad c_{i+1} \leqq s_{i+1}^2/\sigma_{i+1}^2 \leqq d_{i+1}$$

holds for fixed S_i, is a constant depending only on n, i, a_i, c_{i+1}, and d_{i+1}. As a matter of fact,

(27)
$$P_{i+1} = \int_{c_{i+1}}^{d_{i+1}} G_i(a_i^2 x) g_{n-i}(x)\, dx \qquad (i = 1, 2, \cdots, p - 1),$$

where

(28)
$$G_\nu(x) = \int_0^x g_\nu(\xi)\, d\xi$$

and

(29)
$$g_\nu(x) = \frac{e^{-x} x^{\frac{1}{2}\nu - 1}}{2^\nu \Gamma(\frac{1}{2}\nu)}$$

Also, given preassigned positive constants b_1, $c_1 (b_1 < c_1)$, the marginal probability P_1 that

(30)
$$c_1 \leqq s_1^2/\sigma_1^2 \leqq d_1$$

is given by

(31)
$$P_1 = \int_{c_1}^{d_1} g_n(x)\, dx$$

By an argument similar to that which follows (17) in section 2.3, we obtain the probability P that simultaneously

(32)
$$c_i \leqq s_i^2/\sigma_i^2 \leqq d_i \qquad (i = 1, 2, \cdots, p),$$
$$(b_i - \beta_i)' S_i (b_i - \beta_i)/s_{i+1}^2 \leqq a_i^2 \qquad (i = 1, 2, \cdots, p - 1)$$

as

$$P = \prod_{i=1}^{p} P_i .$$

Now, as in Section 2.3, for a given subset T_i of the integers $1, 2, \cdots, i$, writing $T_i[\beta_i]$ and $T_i[b_i]$ for the T_i-norms of β_i and b_i respectively, and writing U_{iT_i} for the T_i-submatrix of S_i^{-1},

(33)
$$s_i^2/d_i \leqq \sigma_i^2 \leqq s_i^2/c_i \qquad \text{for } i = 1, 2, \cdots, p,$$
$$T_i[b_i] - a_i s_{i+1} \lambda_{max}^{1/2}(U_{iT_i}) \leqq T_i[\beta_i] \leqq T_i[b_i] + a_i s_{i+1} \lambda_{max}^{1/2}(U_{iT_i}$$

for all subsets T_i of $(1, 2, \cdots, i)$ and $i = 1, 2, \cdots, p - 1$. The statement (33) thus provides simultaneous confidence-bounds on p parameters of the type σ_i^2 and $(2^p - p)$ parameters of the form $T_i[\beta_i]$ with probability not less than P.

It is to be noted that to set up simultaneous confidence bounds of the type (32), one has to evaluate the integral (27) which is not usually available in tabulated form. Another meaningful procedure, which, incidentally, avoids this difficulty, is to set up separate sets of simultaneous confidence bounds: one on $\sigma_1^2, \cdots, \sigma_p^2$, using the chi-square distribution for s_i^2/σ_i^2, with a preassigned probability and another set on the step-down regressions β_i, using the variance-ratio distribution for $(b_i - \beta_i)'S_i(b_i - \beta_i)/s_{i+1}^2$, and with a probability not less than a preassigned level.

We suggest a slightly different procedure for testing the hypothesis $\mathcal{H}_0$ that Σ has a specified value Σ_0. This hypothesis may be reformulated in terms of the step-down regression-coefficients and residual variances as follows: the hypothesis $\mathcal{H}_0$ is true if and only if each of the hypotheses

$$\mathcal{H}_{i1} : \sigma_i^2 = \sigma_{i0}^2, \qquad i = 1, 2, \cdots, p,$$

$$\mathcal{H}_{i2} : \beta_i = \beta_{i0}, \qquad i = 1, 2, \cdots, p - 1,$$

is true, where σ_{i0}^2, β_{i0} are derived from Σ_0 the same way as σ_i^2, β_i are derived from Σ. The test procedure suggested is:

accept $\mathcal{H}_0$ if

$$(34) \qquad \begin{aligned} c_i \leqq s_i^2/\sigma_{i0}^2 \leqq d_i \qquad & (i = 1, 2, \cdots, p), \\ (b_i - \beta_{i0})'S_i(b_i - \beta_{i0})/\sigma_{i+1,0}^2 \leqq e_i^2 \qquad & (i = 1, 2, \cdots, p - 1); \end{aligned}$$

otherwise reject $\mathcal{H}_0$.

The level of significance α for this procedure is given by

$$(35) \qquad \alpha = 1 - \left\{\prod_{i=1}^{p} P_i'\right\}\left\{\prod_{i=1}^{p-1} P_i''\right\},$$

where

$$P_i' = \int_{c_i}^{d_i} g_{n-i+1}(x)\, dx,$$

$$P_i'' = G_i(e_i^2).$$

For a given α, the c_i, d_i, e_i's are not uniquely determined. The arbitrariness may be removed, for instance, by the further stipulation that

$$P_1' = P_2' = \cdots = P_p' = P_1'' = P_2'' = \cdots = P_{p-1}'' = \beta \text{ (say)}$$

and that (c_i, d_i) are the locally unbiassed partitioning of the 100 $(1 - \beta)$ per cent critical region based on the chi-square distribution with $n - i + 1$ degrees of freedom. With this choice of the constants c_i, d_i, e_i, the test procedure is locally unbiassed.

3.2 *Two variance-covariance matrices.* With two population variance-covariance

matrices $\Sigma^{(1)}$, $\Sigma^{(2)}$ and two matrices of random variables $S^{(1)}$, $S^{(2)}$ distributed independently in Wishart's form with n_1 and n_2 degrees of freedom respectively, so that $S^{(j)}/n_j$ provides an unbiassed estimate for $\Sigma^{(j)}$, we can use the step-down procedure for testing the hypothesis $\mathcal{H}_0$ that the two variance-covariance matrices are identical or, in symbols,

$$\mathcal{H}_0 : \Sigma^{(1)} = \Sigma^{(2)},$$

and also set up simultaneous confidence bounds for parameters measuring deviations from $\mathcal{H}_0$.

Let us introduce the two sets of step-down regression-coefficients and residual variances: $\beta_i^{(j)}$, $\sigma_i^{(j)}$, $b_i^{(j)}$, and $s_i^{(j)}$. The hypothesis $\mathcal{H}_0$ may be reformulated in terms of the step-down parameters as follows: $\mathcal{H}_0$ is true if and only if the hypotheses

(36)
$$\mathcal{H}_{i1} : \sigma_i^{(1)} = \sigma_i^{(2)}, \qquad i = 1, 2, \cdots, p,$$
$$\mathcal{H}_{i2} : \beta_i^{(1)} = \beta_i^{(2)}, \qquad i = 1, 2, \cdots, p-1,$$

are simultaneously true. We may take $\rho_i = \sigma_i^{(1)}/\sigma_i^{(2)}$ and $T_i[\delta_i]$ as measures of deviation from $\mathcal{H}_0$ where $\delta_i = \beta_i^{(1)} - \beta_i^{(2)}$, T_i is a subset of $(1, 2, \cdots, i)$ and $T_i[\delta_i]$ denotes the T_i-norm of δ_i. In this case, it has not been possible to set-up confidence bounds on all these parameters simultaneously. However, one may proceed as follows. Given pre-assigned positive constants c_i, d_i; $c_i < d_i$, and writing

(37)
$$r_i = \left(\frac{n_1 - i + 1}{n_2 - i + 1}\right)^{-1/2} s_i^{(1)}/s_i^{(2)},$$

we find the probability that

(38)
$$r_i^2/d_i \leq \rho_i^2 \leq r_i^2/c_i, \qquad i = 1, 2, \cdots, p,$$

should hold simultaneously is given by

(39)
$$P = \prod_{i=1}^{p} P_i,$$

where

(40)
$$P_i = \int_{c_i}^{d_i} dF_{n_2-i+1}^{n_1-i+1}(x),$$

in which $F_n^m(x)$ stands for the distribution-function of the variance-ratio statistic with m degrees of freedom for the numerator and n degrees of freedom for the denominator. Therefore, (38) provides simultaneous confidence-bounds on ρ_i^2 $(i = 1, 2, \cdots, p)$ with probability P.

Let us now write $\hat{\delta}_i = b_i^{(1)} - b_i^{(2)}$ and note that if $S_i^{(1)}$ and $S_i^{(2)}$ are fixed, $\hat{\delta}_i$ is distributed in an i-variate normal form with expected value δ_i and variance-covariance matrix

$$[\sigma_{i+1}^{(1)}]^2[S_i^{(1)}]^{-1} + [\sigma_{i+1}^{(2)}]^2[S_i^{(2)}]^{-1}$$

distributed independently of $s_{i+1}^{(1)}$ and $s_{i+1}^{(2)}$. If $\mathcal{H}_{i+1,1}$ is true, we have $\sigma_{i+1}^{(1)} = \sigma_{i+1}^{(2)} = \sigma_{i+1}$, say. In that case, if $S_i^{(1)}$ and $S_i^{(2)}$ are fixed, $\hat{\delta}_i$ is distributed in an i-variate normal form with expected value δ_i and dispersion matrix $C_i . \sigma_{i+1}^2$ where

(41) $$C_i = \{S_i^{(1)}\}^{-1} + \{S_i^{(2)}\}^{-1}.$$

Also, $\hat{\delta}_i$ is distributed independently of u_1 and u_2 where

(42) $$u_j = (s_{i+1}^{(j)})^2/\sigma_{i+1}^2 \qquad (j = 1, 2)$$

and u_j is distributed as a chi-square with $(n_j - i)$ degrees of freedom. Consequently, writing

(43) $$s_{i+1}^2 = (s_{i+1}^{(1)})^2 + (s_{i+1}^{(2)})^2$$

we find that if $\mathcal{H}_{i+1,1}$ is true and $S_i^{(j)}$ are fixed $(j = 1, 2)$ the statistics

(44) $$(\hat{\delta}_i - \delta_i)'C_i^{-1}(\hat{\delta}_i - \delta_i)/s_{i+1}^2$$

and

(45) $$\frac{n_2 - i}{n_1 - i}\left(\frac{s_{i+1}^{(1)}}{s_{i+1}^{(2)}}\right)^2$$

are distributed independently as variance-ratios, (44) with i and $(n_1 + n_2 - 2i)$ degrees of freedom, and (45) with $(n_1 - i)$ and $(n_2 - i)$ degrees of freedom.

Therefore, given pre-assigned positive quantities e_i^2 the probability P' that

(46) $$(\hat{\delta}_i - \delta_i)'C_i^{-1}(\hat{\delta}_i - \delta_i)/s_{i+1}^2 \leqq e_i^2, \quad i = 1, 2, \cdots, p - 1,$$

should hold simultaneously is equal to

(47) $$P' = \prod_{i=1}^{p-1} P_i',$$

where

(48) $$P_i' = F_{n_1+n_2-2i}^{i}(e_i^2)$$

provided $\mathcal{H}_{i1}$ is true for $i = 2, 3, \cdots, p$. From (45), we get the following simultaneous confidence-bounds (49) on the T_i-norms of δ_i where T_i is a subset of $(1, 2, \cdots, i)$ (under the highly restrictive condition that $\mathcal{H}_{i1}$ is true) for $i = 2, 3, \cdots, p$:

(49) $$T_i[\hat{\delta}_i] - e_i s_{i+1} \lambda_{\max}^{1/2}(C_{i(T_i)}) \leqq T_i[\delta_i] \leqq T_i[\hat{\delta}_i] + e_i s_{i+1} \lambda_{\max}^{1/2}(C_{i(T_i)})$$

with probability not less than P', where $C_{i(T_i)}$ is the T_i-submatrix of C_i.

To test the hypothesis $\mathcal{H}_0$, the step-down procedure suggested is:

accept $\mathcal{H}_0$ if

(50) $$\begin{aligned} &(\hat{\delta}_i - \delta_i)'C_i^{-1}(\hat{\delta}_i - \delta_i)/s_{i+1}^2 \leqq e_i^2, \quad i = 1, 2 \cdots, p - 1, \\ &c_i \leqq \frac{n_2 - i + 1}{n_1 - i + 1}\frac{s_i^{(1)}}{s_i^{(2)}} \leqq d_i, \qquad i = 1, 2, \cdots, p, \end{aligned}$$

and, otherwise, reject $\mathcal{H}_0$,

where e_i^2, c_i, $d_i(c_i < d_i)$ are pre-assigned positive constants. The level of significance α is given by

$$(51) \qquad \alpha = 1 - \left\{\prod_{i=1}^{p} P_i\right\}\left\{\prod_{i=1}^{p-1} P_i'\right\},$$

where P_i is given by (40) and P_i' by (18). For a pre-assigned value of α, the constants c_i, d_i, e_i^2 are uniquely determined if we stipulate that

$$P_1 = P_2 = \cdots = P_p = P_1' = P_2' = \cdot \cdot = P_{p-1}' = \beta, \quad \text{say},$$

and that (c_i, d_i) gives an unbiassed partitioning of the $100(1 - \beta)$ per cent critical region of the variance-ratio distribution with i and $n_1 + n_2 - 2i$ degrees of freedom. With this choice the step-down test is locally unbiassed.

4. Acknowledgment. The author wishes to thank Professor S. N. Roy for kindly going through the manuscript and suggesting improvements.

REFERENCES

[1] C. R. Rao, "On transformations useful in the distribution problems of least squares," *Sankhyā*, Vol. 12 (1952-53), pp 339-346
[2] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," *Ann. Math. Stat*, Vol. 24 (1953), pp 220-238
[3] S. N. Roy, "A report on some aspects of multivariate analysis," North Carolina Institute of Statistics Mimeograph Series No 121, 1954
[4] S. N. Roy and R. C. Bose, "Simultaneous confidence interval estimation," *Ann Math Stat.*, Vol. 24 (1953), pp 513-536
[5] S. N. Roy and R. E. Bargmann, "Tests of multiple independence and the associated confidence-bounds," North Carolina Institute of Statistics Mimeograph Series No. 175, 1957.
[6] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol 24 (1932), pp. 471-494

THE LIMITING DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT IN THE EXPLOSIVE CASE

By John S. White

Aero Division, Minneapolis Honeywell Regulator Company, Minneapolis Minnesota

1. Introduction and summary. Several authors have studied the discrete stochastic process (x_t) in which the x's are related by the stochastic difference equation

$$x_t = \alpha x_{t-1} + u_t, \qquad t = 1, 2, \cdots, T, \tag{1.1}$$

where the u's are unobservable disturbances, independent and identically distributed with mean zero and variance σ^2, and α is an unknown parameter.

The statistical problem is to find some appropriate function of the x's as an estimator for α and examine its properties.

We may rewrite (1.1) as

$$x_t = u_t + \alpha u_{t-1} + \cdots + \alpha^{t-1} u_1 + \alpha^t x_0. \tag{1.2}$$

From (1.2) we see that the distribution of the successive x's is not uniquely determined by that of the u's alone. The distribution of x_0 must also be specified. Three distributions which have been proposed for x_0 are the following:

(A) x_0 = a constant (with probability one),
(B) x_0 is normally distributed with mean zero and variance $\sigma^2/(1 - \alpha^2)$,
(C) $x_0 = x_T$.

Distribution (B) is perhaps the most appealing from a physical point of view, since if x_0 has this distribution and if the u's are normally distributed, then the process is stationary (e.g., see Koopmans [4]). However, there are several analytic difficulties which arise in the statistical treatment of this process. Distribution (C), the so-called circular distribution, has been proposed as an approximation to (B) and is much easier to analyze (e.g., see Dixon [2]). Distribution (A) has been studied extensively by Mann and Wald [5]. An interesting feature of distribution (A) is that α may assume any finite value, while for distributions (B) and (C) α must be between -1 and 1. From (1.2) we see that a process satisfying (1.1) and (A) has

$$\operatorname{var}(x_t) = \sigma^2(1 + \alpha^2 + \cdots + \alpha^{2(t-1)}). \tag{1.3}$$

If $|\alpha| \geqq 1$, $\lim_{t=\infty} \operatorname{var}(x_t) = \infty$ and the process is said to be "explosive."

Mann and Wald [5] considered only the case $|\alpha| < 1$. They showed that the least squares estimator for α is the serial correlation coefficient[1]

$$\hat{\alpha} = \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2} \tag{1.4}$$

Received December 30, 1955; revised May 27, 1958.

[1] In this paper, the summation sign $\sum$ will always refer to summation from $t = 1$ to $t = T$.

and that (for $|\alpha| < 1$) this estimator is asymptotically normally distributed with mean α and variance $(1 - \alpha^2)/T$. Rubin [6] showed that the estimator $\hat{\alpha}$ is consistent (i.e., plim $\hat{\alpha} = \alpha$) for all α.

In this paper the asymptotic distribution of $\hat{\alpha}$ will be studied under the assumption that the u's are normally distributed. For $|\alpha| > 1$, it is shown that the asymptotic distribution of α is the Cauchy distribution For $|\alpha| = 1$, a moment generating function is found, the inversion of which will yield the asymptotic distribution.

2. The distribution of $\hat{\alpha} - \alpha$. From equation (1.1) and condition (A) the joint distribution of

$$x' = (x_1, x_2, \cdots, x_T)$$

is easily found to be

$$f(x') = \frac{\exp[(-1/2\sigma^2)\sum(x_t - \alpha x_{t-1})^2]}{(2\pi\sigma^2)^{T/2}}. \tag{2.1}$$

The maximum likelihood estimator for α is then the least-squares estimator $\hat{\alpha}$ Since we shall be considering only the distribution of

$$\hat{\alpha} = \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2},$$

we may, without loss of generality, take $\sigma^2 = 1$ For the time being we shall also set $x_0 = 0$.

We may now write (2.1) in matrix form as follows

$$f(x') = \frac{\exp(-\frac{1}{2}x'Px)}{(2\pi)^{T/2}}, \tag{2.2}$$

where P is the $T \times T$ matrix

$$P = \begin{bmatrix} 1+\alpha^2 & -\alpha & 0 & 0 & & & \\ -\alpha & 1+\alpha^2 & -\alpha & 0 & & & \\ 0 & -\alpha & 1+\alpha^2 & -\alpha & & & \\ & & & & \ddots & & \\ & & & & -\alpha & 1+\alpha^2 & -\alpha \\ & & & & 0 & -\alpha & 1 \end{bmatrix} \tag{2.3}$$

Since $\hat{\alpha}$ is a consistent estimator for α, we shall consider the distribution of $\hat{\alpha} - \alpha$ rather than that of α alone. We have

$$\begin{aligned} \hat{\alpha} - \alpha &= \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2} - \alpha \\ &= \frac{\sum x_t x_{t-1} - \alpha \sum x_{t-1}^2}{\sum x_{t-1}^2} \\ &= \frac{x'Ax}{x'Bx}, \end{aligned} \tag{2.4}$$

where A and B are the $T \times T$ matrices

$$A = -\tfrac{1}{2}\begin{bmatrix} 2\alpha & -1 & 0 & & & & \\ -1 & 2\alpha & -1 & & & & \\ 0 & -1 & 2\alpha & & & & \\ & & & \cdots & & & \\ & & & & -1 & 2\alpha & -1 \\ & & & & 0 & -1 & 0 \end{bmatrix},$$

(2.5)

$$B = \begin{bmatrix} 1 & 0 & 0 & & & & \\ 0 & 1 & 0 & & & & \\ 0 & 0 & 1 & & & & \\ & & & \cdots & & & \\ & & & & 0 & 1 & 0 \\ & & & & 0 & 0 & 0 \end{bmatrix}.$$

Let $m(u, v)$ be the joint moment generating function of $x'Ax$ and $x'Bx$. We have

$$\begin{aligned} m(u, v) &= E\,(\exp\{x'Axu + x'Bxv\}) \\ &= (2\pi)^{-T/2} \int \exp\,(x'Axu + x'Bxv - x'Px/2)\, dx \qquad (2.6) \\ &= (2\pi)^{-T/2} \int \exp\,(-x'Dx/2)\, dx, \end{aligned}$$

where D is the $T \times T$ matrix

$$D = P - 2Au - 2Bv = \begin{bmatrix} p & q & 0 & & & & \\ q & p & q & & & & \\ 0 & q & p & & & & \\ & & & \cdots & & & \\ & & & & q & p & q \\ & & & & 0 & q & 1 \end{bmatrix}, \qquad (2.7)$$

$$p = 1 + \alpha^2 - 2v + 2\alpha u, \qquad q = -(\alpha + u).$$

By a well-known integration formula (Cramer [1], Eq. (11.12.2.), p. 120) we have

$$m(u, v) = (2\pi)^{-T/2} \int \exp\left(-\frac{x'Dx}{2}\right) dx = (\det D)^{-\frac{1}{2}}. \qquad (2.8)$$

If we now write $\det D = D(T)$, we note that expanding (2.7) by the elements of the first column gives the difference equation

$$D(T) = pD(T-1) - q^2 D(T-2). \qquad (2.9)$$

From the initial values $D(1) = 1$ and $D(2) = p - q^2$, we obtain

$$D(T) = \frac{1-s}{r-s}\, r^T + \frac{1-r}{s-r}\, s^T, \qquad (2.10)$$

where r and s are roots of the equation $x^2 - px + q^2 = 0$, that is

(2.11) $$r, s = (p \pm \sqrt{p^2 - 4q^2})/2$$

The inversion of $m(u, v) = D(T)^{-\frac{1}{2}}$ seems out of the question for finite T. The inversion of a certain limiting form of $m(u, v)$ will be discussed in Section 4.

3. The standardizing function $g(T)$. Since $\hat{\alpha}$ is consistent the limiting distribution of $\hat{\alpha} - \alpha$ is the unitary distribution. The first problem then is to find some function of T, say $g(T)$, such that the limiting distribution of $g(T)\ (\hat{\alpha} - \alpha)$ is non-degenerate. We note that the results of Mann and Wald (Eq. (1.1) above) give $g(T) = (T/\{1 - \alpha^2\})^{\frac{1}{2}}$ for $|\alpha| < 1$, since $(T/\{1 - \alpha^2\})^{\frac{1}{2}}\ (\hat{\alpha} - \alpha)$ has a limiting normal distribution. The function $g^2(T)$ corresponds roughly to the reciprocal of the asymptotic variance of $(\hat{\alpha} - \alpha)$, or in Fisher's terminology the "information" on α supplied by the sample.

The "information" on α may be obtained explicitly as follows. Let f be the density function (2.1) with $x_0 = 0$ and $\sigma^2 = 1$. The "information," say $I(\alpha)$, is then defined as

$$I(\alpha) = E\left(-\frac{d^2 \log f}{d\alpha^2}\right)$$

$$= E\left(\sum x_{t-1}^2\right)$$

(3.1)
$$= \frac{1}{1 - \alpha^2}\left(T - \frac{1 - \alpha^{2T}}{1 - \alpha^2}\right) \quad \text{if} \quad |\alpha| \neq 1$$

$$= \frac{T(T - 1)}{2} \quad \text{if} \quad |\alpha| = 1$$

If the x's had been independent random variables, then $I(\alpha)\ (\hat{\alpha} - \alpha)$ would be asymptotically $N\ (0, 1)$ (Cramer [1], Eq.(33.3.4), p. 503). This, of course, is not the case. This approach does, however, give an heuristic method for finding a function $g(T)$ such that $g(T)\ (\hat{\alpha} - \alpha)$ has a non-degenerate limiting distribution.

We might now take $g(T) = [I(\alpha)]^{\frac{1}{2}}$; however, it will simplify the computations to use slight modifications which are asymptotically equivalent to $[I(\alpha)]^{\frac{1}{2}}$. We choose

$$g(T) = \sqrt{\frac{T}{1 - \alpha^2}} \qquad \text{for } |\alpha| < 1,$$

(3.2)
$$= \frac{T}{\sqrt{2}} \qquad \text{for } |\alpha| = 1,$$

$$= \frac{|\alpha|^T}{\alpha^2 - 1} \qquad \text{for } |\alpha| > 1$$

In the next section it will be shown that $g(T)\ (\hat{\alpha} - \alpha)$ has a non-degenerate distribution for all values of α.

where A and B are the $T \times T$ matrices

$$A = -\tfrac{1}{2}\begin{bmatrix} 2\alpha & -1 & 0 & & & & \\ -1 & 2\alpha & -1 & & & & \\ 0 & -1 & 2\alpha & & & & \\ & & & \cdots & & & \\ & & & & -1 & 2\alpha & -1 \\ & & & & 0 & -1 & 0 \end{bmatrix},$$

(2.5)

$$B = \begin{bmatrix} 1 & 0 & 0 & & & & \\ 0 & 1 & 0 & & & & \\ 0 & 0 & 1 & & & & \\ & & & \cdots & & & \\ & & & & 0 & 1 & 0 \\ & & & & 0 & 0 & 0 \end{bmatrix}.$$

Let $m(u, v)$ be the joint moment generating function of $x'Ax$ and $x'Bx$. We have

$$\begin{aligned} m(u, v) &= E\,(\exp\{x'Axu + x'Bxv\}) \\ &= (2\pi)^{-T/2} \int \exp\,(x'Axu + x'Bxv - x'Px/2)\,dx \\ &= (2\pi)^{-T/2} \int \exp\,(-x'Dx/2)\,dx, \end{aligned} \tag{2.6}$$

where D is the $T \times T$ matrix

$$D = P - 2Au - 2Bv = \begin{bmatrix} p & q & 0 & & & & \\ q & p & q & & & & \\ 0 & q & p & & & & \\ & & & \cdots & & & \\ & & & & q & p & q \\ & & & & 0 & q & 1 \end{bmatrix}, \tag{2.7}$$

$$p = 1 + \alpha^2 - 2v + 2\alpha u, \qquad q = -(\alpha + u).$$

By a well-known integration formula (Cramer [1], Eq. (11.12.2.), p. 120) we have

$$m(u, v) = (2\pi)^{-T/2} \int \exp\left(-\frac{x'Dx}{2}\right) dx = (\det D)^{-\frac{1}{2}}. \tag{2.8}$$

If we now write $\det D = D(T)$, we note that expanding (2.7) by the elements of the first column gives the difference equation

$$D(T) = pD(T-1) - q^2 D(T-2). \tag{2.9}$$

From the initial values $D(1) = 1$ and $D(2) = p - q^2$, we obtain

$$D(T) = \frac{1-s}{r-s}\, r^T + \frac{1-r}{s-r}\, s^T, \tag{2.10}$$

where r and s are roots of the equation $x^2 - px + q^2 = 0$, that is

$$r, s = (p \pm \sqrt{p^2 - 4q^2})/2 \tag{2.11}$$

The inversion of $m(u, t) = D(T)^{-\frac{1}{2}}$ seems out of the question for finite T. The inversion of a certain limiting form of $m(u, t)$ will be discussed in Section [illegible].

3. The standardizing function $g(T)$. Since $\hat{\alpha}$ is consistent the limiting distribution of $\hat{\alpha} - \alpha$ is the unitary distribution. The first problem then is to find some function of T, say $g(T)$, such that the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$ is non-degenerate. We note that the results of Mann and Wald [illegible] give $g(T) = (T/|1 - \alpha^2|)^{\frac{1}{2}}$ for $|\alpha| < 1$, since $(T/(1 - \alpha^2))^{\frac{1}{2}}(\hat{\alpha} - \alpha)$ has a limiting normal distribution. The function $g(T)$ corresponds roughly to the reciprocal of the asymptotic variance of $(\hat{\alpha} - \alpha)$ [illegible] the "information" on α supplied by the sample.

The "information" on α may be obtained [illegible]. Let f be the density function (2.1) with $x_0 = 0$ and $\sigma^2 = 1$. The "information," say $I(\alpha)$, is then defined as

$$\begin{aligned} I(\alpha) &= E\left(-\frac{d^2 \log f}{d\alpha^2}\right) \\ &= E\left(\sum x_{t-1}^2\right) \\ &= \frac{1}{1 - \alpha^2}\left(T - \frac{1 - \alpha^{2T}}{1 - \alpha^2}\right) \quad \text{if} \quad |\alpha| \neq 1 \\ &= \frac{T(T - 1)}{2} \quad \text{if} \quad |\alpha| = 1 \end{aligned} \tag{3.1}$$

If the x's had been independent random variables, then $I(\alpha)^{\frac{1}{2}}(\hat{\alpha} - \alpha)$ would be asymptotically $N(0, 1)$ (Cramér [illegible]). That of course is not the case. This approach does, however, [illegible] function $g(T)$ such that $g(T)(\hat{\alpha} - \alpha)$ has a [illegible].

We might now take $g(T) = I(\alpha)^{\frac{1}{2}}$, however [illegible] to use slight modifications which are asymptotically [illegible] choose

$$\begin{aligned} g(T) &= [illegible] \quad \text{for} \quad |\alpha| < 1 \\ &= [illegible] \quad \text{for} \quad |\alpha| = 1 \\ &= [illegible] \quad \text{for} \quad |\alpha| > 1 \end{aligned} \tag{3.2}$$

In the next section [illegible] $g(T)(\hat{\alpha} - \alpha)$ [illegible] distribution [illegible]

4. The limiting distribution of $g(T)(\hat{\alpha} - \alpha)$. We shall first consider the joint distribution of $x'Ax/g(T)$ and $x'Bx/g^2(T)$. Let $M(U, V)$ be the joint moment generating function of these two statistics. We then have

$$\begin{aligned} M(U, V) &= E[\exp x'AxU/g(T) + x'BxV/g^2(T)] \\ &= m[U/g(T), V/g^2(T)], \end{aligned} \tag{4.1}$$

where $m(u, v)$ is the joint moment generating function (2.6).

From (2.10) and (2.11) with $g = g(T)$, $u = U/g$ and $v = V/g^2$, we have

$$\begin{aligned} M(U, V) &= D(T)^{-\frac{1}{2}} \\ &= \frac{1 - s}{r - s} r^T + \frac{1 - r}{s - r} s^T, \end{aligned} \tag{4.2}$$

$$\begin{aligned} r, s = \tfrac{1}{2}[1 + \alpha^2 + 2\alpha U/g - 2V/g^2 \pm \{(1 - \alpha^2)^2 - 4\alpha(1 - \alpha^2)U/g \\ - 4(1 - \alpha^2)U^2/g^2 - 4(1 + \alpha^2)V/g^2 - 8\alpha UV/g^3 + 4V^2/g^4\}^{1/2}]. \end{aligned} \tag{4.3}$$

For sufficiently large T and $|\alpha| \neq 1$, we may factor $(1 - \alpha^2)$ out of the radical in (4.3) and expand the remaining radical by the binomial theorem. We then have, up to terms of order $O(g^{-3})$

$$\begin{aligned} r, s = \frac{1}{2}\Big[& 1 + \alpha^2 + 2\alpha U/g - 2V/g^2 \\ & \pm \left\{1 - \alpha^2 - 2\alpha U/g - \frac{2(1 + \alpha^2)V}{(1 - \alpha^2)g^2} - \frac{2U^2}{(1 - \alpha^2)g^2} + O(g^{-3})\right\}\Big]. \end{aligned} \tag{4.4}$$

Taking r with the plus sign and s with the minus sign we have

$$\begin{aligned} r &= 1 - \frac{U^2 + 2V}{(1 - \alpha^2)g^2} + O(g^{-3}), \\ s &= \alpha^2 + 2\alpha U/g + \frac{U^2 + 2\alpha^2 V}{(1 - \alpha^2)g^2} + O(g^{-3}). \end{aligned} \tag{4.5}$$

Substituting the appropriate values of $g(T)$ from (3.2), we have

$$\begin{aligned} r &= 1 - \frac{U^2 + 2V}{T} + O(T^{-\frac{3}{2}}) \qquad \text{for } |\alpha| < 1, \\ s &= \alpha^2 + 2\alpha\sqrt{\frac{1 - \alpha^2}{T}}\, U + \frac{U^2 + 2\alpha^2 V}{T} + O(T^{-\frac{3}{2}}). \end{aligned} \tag{4.6}$$

$$\begin{aligned} r &= 1 + \frac{(U^2 + 2V)(\alpha^2 - 1)}{\alpha^{2T}} + O(|\alpha|^{-3T}) \qquad \text{for } |\alpha| > 1, \\ s &= \alpha^2 + \frac{2\alpha U(\alpha^2 - 1)}{|\alpha|^T} - \frac{(U^2 + 2\alpha^2 V)(\alpha^2 - 1)}{\alpha^{2T}} + O(|\alpha|^{-3T}). \end{aligned} \tag{4.7}$$

If $|\alpha| = 1$, the expansion in (4.4) is not valid, however, from (4.3), we have

$$r = 1 + \frac{\sqrt{2}\alpha U}{T} + \frac{2i\sqrt{V}}{T} + O(T^{-2}) \qquad \text{for } |\alpha| = 1,$$

(4.8)

$$s = 1 + \frac{\sqrt{2}\alpha U}{T} - \frac{2i\sqrt{V}}{T} + O(T^{-2})$$

Substituting these results in (4.2), we have

$$\lim M(U, V) = \exp (V + U^2/2) \qquad \text{for } |\alpha| < 1,$$

$$= (1 - U^2 - 2V)^{-1/2} \qquad \text{for } |\alpha| > 1,$$

(4.9)

$$= \exp (\sqrt{2}\alpha U) \left(\cos 2\sqrt{V} - \frac{\sqrt{2}\alpha U}{2\sqrt{V}} \sin 2\sqrt{V}\right)^{-1/2}$$

$$\text{for } |\alpha| = 1$$

The next problem is to obtain the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$ from $\lim M(U, V)$. Since $g(T)(\hat{\alpha} - \alpha) = g(T)x'Ax/x'bX$, the problem is one of finding the distribution of the ratio of two random variables. One method of solution has been proposed by Gurland [3]. Let X and Y be two random variables, Prob $(Y > 0) = 1$. We wish to determine the distribution of $Z = X/Y$. Let $W = W_z = X - zY$. Then we have

$$\text{Prob } (Z < z) = \text{Prob } (X/Y < z)$$

(4.10)

$$= \text{Prob } (X - zY < 0)$$

$$= \text{Prob } (W_z < 0).$$

If the distribution of W can be found, the distribution of Z will immediately follow. Frequently the distribution of W can be found from that of X and Y by means of moment generating functions. Let

(4.11) $\qquad m(w) = E(\exp\{Ww\}), \qquad m^*(u, v) = E(\exp\{Xu + Yv\}),$

then

$$m(w) = E(\exp\{X - zY\}w) = E(\exp\{Xw - Yzw\}) = m^*(w, -zw)$$

To apply this technique to the problem at hand, we set $W = x'Ax/g - zx'Bx/g^2$. From (4.1), (4.2) and (4.9) we have

$$m(w) = M(w, -zw),$$

$$\lim m(w) = \exp (-zw + w^2/2) \qquad \text{for } |\alpha| < 1,$$

(4.12)

$$= (1 + 2zw - w^2)^{-1/2} \qquad \text{for } |\alpha| > 1,$$

$$= \left\{\exp (\sqrt{2}\alpha w) \left(\cos 2\sqrt{-zw} - \frac{\sqrt{2}\alpha w}{2\sqrt{-zw}} \sin 2\sqrt{-zw}\right)\right\}^{-1/2}$$

$$\text{for } |\alpha| = 1$$

The inversion of lim $m(w)$ is trivial for $|\alpha| < 1$. The moment generating function $\exp(-zw+w^2/2)$ is immediately recognized as that of a random variable which is normally distributed with mean $-z$ and variance 1. Hence we have

$$\text{(4.13)}\quad \begin{aligned} \lim \operatorname{Prob}(W < 0) &= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{0} \exp(-\{t+z\}^2/2)\,dt \\ &= (2\pi)^{-1/2} \int_{-\infty}^{z} \exp(-t^2/2)\,dt \\ &= \lim \operatorname{Prob}\{g(T)(\hat{\alpha}-\alpha) < z\}, \end{aligned}$$

i.e., $g(T)(\hat{\alpha}-\alpha)$ is asymptotically normal with mean 0 and variance 1.

For $|\alpha| > 1$, the inverse of lim $m(w)$ might be obtained directly in terms of Bessel functions; however, it is more appealing from a statistical point of view to proceed as follows. Let X and Y be independent chi-squared variables with one degree of freedom. Then $E(\exp\{Xw\}) = E(\exp\{Yw\}) = (1-2w)^{-1/2}$ is their common moment generating function. Now set $R = aX - bY$, the moment generating function of R will be

$$\text{(4.14)}\quad \begin{aligned} m_R(w) = E(\exp\{Rw\}) &= E(\exp\{aX - bY\}w) \\ &= (\{1-2aw\}\{1+2bw\})^{-1/2}. \end{aligned}$$

In particular if we set

$$\text{(4.15)}\quad 2a = \sqrt{1+z^2} - z, \qquad 2b = \sqrt{1+z^2} + z,$$

we have

$$\text{(4.16)}\quad m_R(w) = (1 + 2zw - w^2)^{1/2} = \lim m(w).$$

Hence, the limiting distribution of W, for $|\alpha| > 1$, is the same as the distribution of $R = aX - bY$. We then have

$$\text{(4.17)}\quad \begin{aligned} \lim \operatorname{Prob}(W < 0) &= \operatorname{Prob}(aX - bY < 0) \\ &= \operatorname{Prob}(X < bY/a) \\ &= \frac{1}{2\pi}\int_0^{\infty}\int_0^{by/a} \frac{\exp(-x/2 - y/2)}{\sqrt{xy}}\,dx\,dy \\ &= \lim \operatorname{Prob}\{g(T)(\hat{\alpha}-\alpha) < z\} = \text{say } F(z). \end{aligned}$$

The density function corresponding to $F(z)$ is

$$\text{(4.18)}\quad \begin{aligned} f(z) = \frac{dF(z)}{dz} &= \frac{1}{2\pi}\int_0^{\infty} \sqrt{a/b}\exp(-by/2a - y/2)\left\{\frac{d(b/a)}{dz}\right\}dy \\ &= \frac{1}{2\pi}\sqrt{a/b}\,\frac{2}{1+(b/a)}\,\frac{d(b/a)}{dz} \\ &= \frac{1}{\pi}\,\frac{1}{1+z^2} \qquad \text{(by (4.15))}. \end{aligned}$$

Hence the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$, for $|\alpha| > 1$, is the Cauchy distribution.

We have been unable to invert $\lim m(w)$ when $|\alpha| = 1$. In the next section certain results concerning this limit and more general problems of this type will be discussed.

If we now let $x_0 = c$, a non-zero constant, the analysis proceeds much as before. Let A, B, P, and D be the $T \times T$ matrices defined in (2.3), (2.5) and (2.7). We then have, analogous to (2.1) and (2.4),

$$f(x') = (2\pi)^{-T/2} \exp(cx_1\alpha - \alpha^2c^2/2 - x'Px/2),$$

(4.19)
$$\hat{\alpha} - \alpha = \frac{x'Ax + cx_1 - \alpha c^2}{x'Bx + c^2}.$$

The joint moment generating function of $x'Ax + cx_1 - \alpha^2c^2$ and $x'Bx + c^2$ is

(4.20)
$$\begin{aligned} m(u, v) &= E\left(\exp\{(x'Ax + cx_1 - \alpha c^2)u + (x'Bx + c^2)v\}\right) \\ &= \left\{\exp\left(c^2v - c^2\alpha u - \frac{\alpha^2c^2}{2}\right)\right\}(2\pi)^{-T/2} \\ &\qquad \cdot \int \exp\left([u + \alpha]cx_1 - \frac{x'Dx}{2}\right) dx \\ &= \exp\left(c^2v - c^2\,\alpha u - \frac{\alpha^2c^2}{2}\right)\exp\left\{(u + \alpha)^2\frac{c^2}{2}\,\frac{D(T - 1)}{D(T)}\right\} D(T)^{-1/2}. \end{aligned}$$

(4.21)
$$\begin{aligned} \lim m(U/g, V/g^2) &= \lim M(U, V) \\ &= \lim\left\{D(T)^{-1}\exp\left(\frac{-\alpha^2c^2}{2}\left[1 - \frac{D'(T - 1)}{D(T)}\right]\right)\right\}, \end{aligned}$$

where $D(T)$ is as defined in (4.2) while $D'(T - 1)$ is defined in a similar fashion but with $g = g(T)$.

For $|\alpha| \leqq 1$, it follows from (4.6) and (4.8) that, since $g(T)$ and $g(T - 1)$ are of the same order,

$$\lim D(T) = \lim D'(T - 1)$$

and hence

(4.22)
$$\lim m(U/g, V/g^2) = \lim M(U, V) = \lim D(T)^{-1/2}$$

We see that this limit is the same as that for $x_0 = 0$ as given in (4.9) and hence the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$ does not depend on the initial value x_0 for $|\alpha| \leqq 1$.

For $|\alpha| > 1$ we have, from (4.7),

(4.23)
$$\begin{aligned} \lim D(T) &= 1 - (U + 2V), \\ \lim D'(T - 1) &= \frac{(U + 2V)}{\alpha^2}, \end{aligned}$$

and in place of (4.22) we have

$$\text{(4.24)}\quad \begin{aligned} \lim M(U, V) &= \lim D(T)^{-1/2} \exp\left(-\frac{\alpha^2 c^2}{2}\left[1 - \frac{D'(T-1)}{D(T)}\right]\right) \\ &= (1 - U^2 - 2V)^{-1/2} \exp\left\{\frac{(\alpha^2 - 1)c^2}{2}\left(\frac{U^2 + 2V}{1 - U^2 - 2V}\right)\right\} \end{aligned}$$

This moment generating function may be inverted by the methods of Section to give

$$\text{(4.25)}\quad f(x) = \frac{e^{-q}}{\sqrt{\pi}(1 + x^2)} \sum_{k=0}^{\infty} \left(\frac{q}{1 + x^2}\right)^k \frac{1}{\Gamma(k + \frac{1}{2})}, \qquad q = \frac{c^2(\alpha^2 - 1)}{2},$$

as the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$. We note that for $c = 0$, $f(x)$ is th Cauchy distribution as obtained in (4.18).

6. Final remarks. The results of Mann and Wald [5] show that the limitin distribution of $g(T)(\hat{\alpha} - \alpha)$, for $|\alpha| < 1$, is also $N(0, 1)$ if, rather than assumin that the "errors" u_t are normally distributed, we merely assume that all of th moments of the u's are finite. This is another example of an invariance principl which seems to hold quite generally for the limiting distributions of function o random variables. Roughly speaking, there seems to be an unproved (and un stated) theorem that the limiting distribution of a function of a sequence o independent random variables, with suitable restrictions on these random vari ables, depends only on the form of the function and is the same as the distributio of a related functional on a stochastic process.

A general result of this form is Donsker's Theorem [7] which gives the limiting distribution of any function of sums of independent identically distribute random variables with finite variances as the distribution of a correspondin functional on the Wiener process. It is conjectured that this type of reasonin will show that the results of Mann and Wald will still hold if the u's are merely assumed to have finite variances.

For $\alpha = 1$, application of Donsker's Theorem shows that the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$ is the same as the distribution of the functional

$$G[x(\cdot)] = \frac{\int_0^1 x(t)\, dx(t)}{\int_0^1 x^2(t)\, dt} = \frac{\frac{1}{2}x^2(1) - \frac{1}{2}}{\int_0^1 x^2(t)\, dt}$$

on the Wiener process, independent of the distribution of the u's. This distribution will be considered in a future paper.

REFERENCES

[1] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[2] W. J. Dixon, "Further contributions to the problem of serial correlation," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 119–144.

[3] J. GURLAND, "Inversion formulae for the distribution of ratios," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 228-237.

[4] T. KOOPMANS, "Serial correlation and quadratic forms in normal variables," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 1–42.

[5] H. B. MANN AND A. WALD, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, Vol. 11 (1943), pp 173-200.

[6] H. RUBIN, "Consistency of maximum likelihood estimates in the explosive case," *Statistical Inference in Dynamic Economic Models*, T Koopmans (ed), John Wiley & Sons, 1950.

[7] M. D. DONSKER, "An invariance principle for certain probability limit theorems," *Mem. Am. Math. Soc.*, No 6, 1951, pp 1 12

A LIMIT THEOREM FOR THE PERIODOGRAM

By Salomon Bochner[1] and Tatsuo Kawata[2]

1. Introduction. Let $\mathcal{E}(t)$ be a real stationary process in the wide sense with mean 0 and let its covariance function and spectral function be $\rho(u)$, $F(x)$ respectively. We assume that $F(x)$ is absolutely continuous and has a spectral density function $p(x)$. The second-named author, [1], has discussed the periodogram

$$(1.1)\qquad J(T) = \frac{1}{4\pi T}\left|\int_{-T}^{T} \mathcal{E}(t)e^{-i\xi t}\,dt\right|^2,$$

in case $\mathcal{E}(t)$ is stationary even of the fourth order, so that the expectation

$$E\mathcal{E}(t)\mathcal{E}(t+u)\mathcal{E}(t+v)\mathcal{E}(t+w) = P(u, v, w)$$

exists and is a function of u, v, w alone. It was also assumed that the function $Q(u, v, w)$, which is the difference between $P(u, v, w)$ and the corresponding fourth moment of a stationary Gaussian process, is the Fourier transform of a function and that the latter function satisfies the Lipschitz condition. Under these assumptions it has proven that (1.1) does not converge in mean to any random variable as $T \to \infty$, but that the covariance function of $J(T)$ and $J(T')$ does tend to a limit whenever T and T' both tend to infinity in a certain related manner, and the limiting value of the covariance function was determined.

The paper involved a rather troublesome manipulation of a Fourier integral, but we have found since that under somewhat different assumptions the complications can be reduced appreciably. In a separate publication, [2], a certain integral transformation was investigated on its own merit, and in the present paper an application of the somewhat modified approach will be made to the problem of the periodogram. The expression (1.1) will be replaced by a more general one, and as regards the difference function $Q(u, v, w)$ the assumptions will be modified as follows. We add expressly the requirement that $Q(u, v, w)$ shall be integrable in E_3, but the requirement that its Fourier transform shall satisfy the Lipschitz condition is being omitted entirely.

2. The Theorem. We shall consider the random variable

$$(2.1)\qquad S(T) = \frac{1}{T}\left|\int_{-\infty}^{\infty} \mathcal{E}(t)M\left(\frac{t}{T}\right)e^{-i\xi t}\right|^2 dt$$

in place of (1.1). We shall call (2.1) a generalized periodogram of $\mathcal{E}(t)$.

Let us assume that

$$(2.2)\qquad P(s_1, s_2, s_3) = Q(s_1, s_2, s_3) + P_G(s_1, s_2, s_3),$$

Received December 6, 1957.

[1] Research done in connection with U. S. Army Ordnance, Office of Ordnance Research, Contract No. D.A.—36-034—ORD-2001.

[2] Work done while the second author was a research fellow at Princeton University supported by the Rockefeller Foundation.

where

$$(2.3)\quad P_G(s_1, s_2, s_3) = \rho(s_1)\,\rho(s_2 - s_3) + \rho(s_2)\rho(s_3 - s_1) + \rho(s_3)\rho(s_1 - s_2).$$

If $\mathcal{E}(t)$ is a stationary Gaussian process, then $Q(s_1, s_2, s_3) \equiv 0$. This assumption was set up first by Magness [3]; see also Parzen [4]

We assume further that

$$(2.4)\qquad Q(s_1, s_2, s_3) \in L_1(E_3),$$

and that the Fourier transform of $Q(s_1, s_2, s_3)$ is also in $L_1(E_3)$, so that

$$(2.5)\qquad q(x_1, x_2, x_3) = \int_{E_3} e^{i(s \cdot x)} Q(s_1, s_2, s_3)\, dv_s,$$

$$(2.6)\qquad Q(s_1, s_2, s_3) = (2\pi)^{-3} \int_{E_3} e^{-i(s \cdot x)} q(x_1, x_2, x_3)\, dv_x,$$

where E_k denotes the whole Euclidean space of k dimension and $(s \cdot x) = s_1x_1 + s_2x_2 + s_3x_3$.

Under these conditions, we obtain the following theorem

THEOREM. *Let $M(\alpha)$ be bounded and integrable in $(-\infty, \infty)$ and let the Fourier transform $K(x)$ of $M(\alpha)$*

$$K(x) = \int e^{ix\alpha} M(\alpha)\, d\alpha$$

satisfy

$$(2.7)\qquad K(x) = O(|x|^{-1}), \text{ as } x \to \infty.$$

Then we have, as T_1 and T_2 tend to infinity such that $T_1/T_2 \to \mu$, $\mu \neq 0$,

$$(2.8)\quad \lim \operatorname{cov}\{S(T_1), S(T_2)\} = \begin{cases} (2\pi)^2\,(|C_\mu^{(1)}|^2 + |C_\mu^{(2)}|^2)p^2(0), & \text{if } \xi = 0, \\ (2\pi)^2\,|C_\mu^{(2)}|^2 p^2(\xi), & \text{if } \xi \neq 0, \end{cases}$$

and

$$(2.9)\quad \lim E\{S(T_1) - S(T_2)\}^2 = \begin{cases} 2(2\pi)^2\,(|C_1^{(1)}|^2 + |C_1^{(2)}|^2 - |C_\mu^{(1)}|^2 \\ \quad - |C_\mu^{(2)}|^2)p^2(0), & \text{if } \xi = 0, \\ 2(2\pi)^2\,(|C_1^{(2)}|^2 - |C_\mu^{(2)}|^2)p^2(\xi), & \text{if } (\xi) \neq 0, \end{cases}$$

provided that $p(x)$ is continuous at ξ, and the constants $C_\mu^{(j)}$ $(j = 1, 2)$ are given by

$$C_\mu^{(1)} = \mu^{\frac{1}{2}} \int_{-\infty}^{\infty} M(\alpha) M(\mu\alpha)\, d\alpha,$$

$$C_\mu^{(2)} = \mu^{\frac{1}{2}} \int_{-\infty}^{\infty} M(\alpha) \bar{M}(\mu\alpha)\, d\alpha.$$

We add a remark. If $\mu \to \infty$, or $\mu \to 0$, then $C_\mu^{(1)}, C_\mu^{(2)}$ converge to 0. This is easily seen from the fact that $C_\mu^{(1)} = C_{1/\mu}^{(1)}$, $C_\mu^{(2)} = \bar{C}_{1/\mu}^{(2)}$, and $|C_\mu^{(j)}| \leq \mu^{1/2} M \int_{-\infty}^{\infty} |M(\alpha)|\, d\alpha \to 0$, $(\mu \to 0)$, M being an upper bound of $M(\alpha)$.

We also note that the theorem implies that the constant

$$|C_1^{(1)}|^2 + |C_1^{(2)}|^2 - |C_\mu^{(1)}|^2 - |C_\mu^{(2)}|^2$$

must be non-negative. This can also be established directly by verifying that it is the value of the double integral

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\,|A(\alpha,\beta)|^2 + (A(\alpha,\beta))^2]\,d\alpha\,d\beta,$$

where

$$A(\alpha,\beta) = M(\alpha)\bar{M}(\beta) - \mu M(\mu\alpha)\bar{M}(\mu\beta).$$

For the proof of the theorem, we first of all state as a lemma, a theorem given in [2].

LEMMA 1. *Let $M_j(\alpha)$ $(j = 0, 1, \cdots, k)$ be bounded and integrable over $(-\infty, \infty)$ and let their Fourier transforms be*

$$K_j(x) = \int_{-\infty}^{\infty} e^{i\alpha x} M_j(\alpha)\,d\alpha, \qquad (j = 0, 1, \cdots, k).$$

Put

$$K(x_1, x_2, \cdots, x_k; T_0, T_1, \cdots, T_k) = \prod_{j=0}^{k} T_j \cdot K(T_0(x_1 + x_2 + \cdots + x_k)) \prod_{j=1}^{k} K_j(T_j x_j)$$

for any positive numbers $T_0, T_1, \cdots, T_k$. Then we have

$$\lim_{T_i\to\infty}\frac{1}{T_0}\int_{E_k} f(x_1, x_2, \cdots, x_k)K(x_1, x_2, \cdots, x_k; T_0, T_1, \cdots, T_k)\,dv_x = C_k(2\pi)^k f(0, \cdots, 0), \qquad (i = 0, 1, \cdots, k),$$

if T_j go to infinity such that $T_0/T_j \to \mu_j$ and $\mu_j \neq 0$ $(j = 1, 2, \cdots, k)$ and $f(x_1, \cdots, x_k)$ satisfies the conditions that the function $f(x_1, \cdots, x_k)$ is continuous and belongs to $L_1(E_k)$ and its Fourier transform

$$g(\alpha_1, \alpha_2, \cdots, \alpha_k) = \int_{E_k} e^{i(\alpha\cdot x)} f(x_1, \cdots, x_k)\,dv_x$$

likewise belongs to $L_1(E_k)$. C_k is

$$C_k = \int_{-\infty}^{\infty} M_0(\alpha)\prod_{j=1}^{k} M_j(-\mu_j\alpha)\,d\alpha.$$

3. A lemma. For the proof of the theorem, we need one more lemma.

LEMMA 2. *Let $K_j(x)$, $(j = 1, 2)$ be a bounded function which is the Fourier transform of a bounded and integrable function $M_j(\alpha)$*

$$K_j(x) = \int_{-\infty}^{\infty} M_j(\alpha)e^{ix\alpha}\,d\alpha, \qquad j = 1, 2, \tag{3.1}$$

and let us assume that

(3.2) $$K_j(x) = O(|x|^{-1}), \qquad \text{as } x \to \infty, \qquad j = 1, 2.$$

(i) If $p(x) \varepsilon L_1(-\infty, \infty)$ and continuous at $-\xi$, then

(3.3) $$(T_1 T_2)^{\frac{1}{2}} \int_{-\infty}^{\infty} K_1(T_1(x + \xi))K_2(T_2(x + \xi))p(x)\, dx$$

converges to

$$2\pi\ C_\mu \cdot p(-\xi),$$

when $T_1, T_2 \to \infty$ *and* $T_1/T_2 \to \mu$ *and* $\mu \neq 0$, *where*

$$C_\mu = \mu^{\frac{1}{2}} \int_{-\infty}^{\infty} M_1(\beta)M_2(\mu\beta)\, d\beta.$$

(ii) *If* $p(x) \varepsilon L_1(-\infty, \infty)$, *and* $p(x)$ *continuous as* $-\xi_1$ *then*

(3.4) $$(T_1 T_2)^{\frac{1}{2}} \int_{-\infty}^{\infty} K_1\{T_1(x + \xi_1)\}K_2\{T_2(x + \xi_2)\}p(x)\, dx$$

converges to zero when $T_1, T_2 \to \infty$ *such that* $T_1/T_2 \to \mu$ *and* $\mu \neq 0$ *and* $\xi_1 \neq \xi_2$.

PROOF. (i) We consider the integral

(3.5) $$(T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1(T_1 x)K_2(T_2 x)\, dx,$$

which is absolutely convergent because K_1, K_2 are bounded and satisfy (3.2) By the Parseval theorem, since $K_j(x) \varepsilon L_2(-\infty, \infty)$, we have

$$(T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1(T_1 x)K_2(T_2 x)\, dx = \frac{2\pi}{(T_1 T_2)^{1/2}} \int_{-\infty}^{\infty} M_1\left(\frac{\alpha}{T_1}\right) M_2\left(\frac{-\alpha}{T_2}\right) d\alpha$$

$$= 2\pi(T_1/T_2)^{1/2} \int_{-\infty}^{\infty} M_1(\beta)M_2\left(-\frac{T_1}{T_2}\beta\right) d\beta.$$

This converges to

$$2\pi C_\mu = 2\pi\mu^{1/2} \int_{-\infty}^{\infty} M_1(\beta)M_2(-\mu\beta)\, d\beta,$$

as is easily seen from the fact that

$$\int_{-\infty}^{\infty} |M_2(a\beta) - M_2(a_0 \beta)|\, d\beta \to 0,$$

if $a \to a_0$ and $a_0 \neq 0$.

Hence it suffices to show that

(3.6) $$I = (T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1(T_1 x)K_2(T_2 x)\{p(x - \xi) - p(-\xi)\}\, dx$$

converges to zero.

We divide I into two parts:

$$I = \int_{|x|<\delta} + \int_{|x|>\delta} = I_1 + I_2,$$

where δ is taken so that $| p(x - \xi) - p(-\xi) | < \epsilon$, for $| x | < \delta$, ϵ being any assigned positive number.

We have

$$|I_1| \leqq \epsilon(T_1 T_2)^{1/2} \int_{|x|<\delta} | K_1(T_1 x)K_2(T_2 x) | \, dx$$

(3.7) $$\leqq \epsilon(T_2/T_1)^{1/2} \int_{-\infty}^{\infty} \left| K_1(u)K_2\left(\frac{T_2}{T_1} u\right)\right| du$$

$$\leqq \epsilon C \int_{-\infty}^{\infty} \frac{du}{1 + u^2},$$

for some constant C, as follows from (3.2).

Next we have

$$|I_2| \leqq (T_1 T_2)^{1/2} \int_{|x|>\delta} | K_1(T_1 x)K_2(T_2 x) | \, |p(x - \xi) \, dx$$

$$+ | p(\xi) | (T_1 T_2)^{1/2} \int_{|x|>\delta} | K_1(T_1 x)K_2(T_2 x) | \, dx$$

$$\leqq \frac{C}{(T_1 T_2)^{1/2}} \int_{|x|>\delta} \frac{| p(x - \xi) |}{x^2} \, dx + \frac{C \, | p(\xi) |}{(T_1 T_2)^{1/2}} \int_{|x|>\delta} \frac{dx}{x^2},$$

for some constant C. Hence we get

(3.8) $$I_2 = o(1)$$

as $T_1T_2 \to \infty$, and this together with (3.7) proves (i).

We shall now prove (ii). We have

$$(T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1(T_1(x + \xi_1))K_2(T_2(x + \xi_2)) \, dx$$

(3.9) $$= \frac{2\pi}{(T_1 T_2)^{1/2}} \int_{-\infty}^{\infty} M_1\left(\frac{\alpha}{T_1}\right) M_2\left(\frac{-\alpha}{T_2}\right) e^{i\alpha(\xi_1 - \xi_2)} \, d\alpha$$

$$= 2\pi(T_1/T_2)^{1/2} \int_{-\infty}^{\infty} M_1(\beta)M_2\left(-\frac{T_2}{T_1}\beta\right) e^{iT_1\beta(\xi_1-\xi_2)} \, d\beta$$

and the difference between this and the expression

(3.10) $$2\pi(T_1/T_2)^{1/2} \int_{-\infty}^{\infty} M_1(\beta)M_2(-\mu\beta)e^{iT_1\beta(\xi_1-\xi_2)} \, d\beta$$

is in absolute value

$$\leqq 2\pi(T_1/T_2)^{1/2} \int_{-\infty}^{\infty} | M_1(\beta) | \left| M_2\left(-\frac{T_1}{T_2}\beta\right) - M_2(-\mu\beta)\right| d\beta$$

$$\leqq C \int_{-\infty}^{\infty} \left| M_2\left(-\frac{T_1}{T_2}\beta\right) - M_2(-\mu\beta)\right| d\beta.$$

But this is as small as we please, for T_1, T_2 large and T_1/T_2 near to μ, provided $\mu \neq 0$.

Now (3.10) tends to zero by Riemann-Lebesgue lemma, and we conclude that (3.9) tends to zero also.

It suffices, then, to show that

$$(3.11) \quad J = (T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1[T_1(x + \xi_1)] K_2[T_2(x + \xi_2)] \{p(x) - p(-\xi_1)\} \, dx$$

converges to zero.

We have

$$(3.12) \quad \begin{aligned} J &= (T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1(T_1 y) K_2\{T_2[y - (\xi_1 - \xi_2)]\} \{p(y - \xi_1) - p(-\xi_1)\} \, dy \\ &= (T_1 T_2)^{1/2} \int_{|y| < \delta} + (T_1 T_2)^{1/2} \int_{|y| > \delta} = J_1 + J_2, \end{aligned}$$

say. Here δ is so chosen that

$$(3.13) \quad |p(y - \xi_1) - p(-\xi_1)| < \epsilon,$$

for $|y| < \delta$ and

$$(3.14) \quad |\xi_1 - \xi_2| - \delta > c > 0,$$

for some positive constant c. Then

$$(3.15) \quad \begin{aligned} |J_1| &\leqq (T_1 T_2)^{1/2} \cdot \epsilon \int_{|y| < \delta} |K_1(T_1 y) K_2[T_2 y - T_2(\xi_1 - \xi_2)]| \, dy \\ &\leqq \epsilon (T_1 T_2)^{1/2} C \int_{|y| < \delta} \frac{dy}{T_2(|\xi_1 - \xi_2| - y)} \\ &\leqq \epsilon (T_1/T_2)^{1/2} C \cdot c \cdot \delta \leqq C\epsilon, \end{aligned}$$

for some constant C by (3.13) and (3.14).

Next we shall consider J_2. We divide J_2 further into two parts,

$$\begin{aligned} J_2 &= (T_1 T_2)^{1/2} \int_{|y| > \delta, |y - (\xi_1 - \xi_2)| > \eta} + (T_1 T_2)^{1/2} \int_{|y| > \delta, |y - (\xi_1 - \xi_2)| < \eta} \\ &= J_{21} + J_{22}, \end{aligned}$$

say, where $0 < \eta < \frac{1}{2} |\xi_1 - \xi_2|$. Then

$$(3.16) \quad \begin{aligned} |J_{21}| &\leqq (T_1 T_2)^{1/2} C \int_{|y| > \delta, |y - (\xi_1 - \xi_2)| > \eta} \frac{1}{T_1 y} \\ &\qquad \cdot \frac{1}{T_2 |y - (\xi_1 - \xi_2)|} (|p(y - \xi_1)| + |p(-\xi_1)|) \, dy \\ &\leqq \frac{C}{(T_1 T_2)^{1/2} \delta \eta} \int_{|y| > \delta, |y - (\xi_1 - \xi_2)| > \eta} \frac{|p(y - \xi_1)| + |p(-\xi_1)|}{y |y - (\xi_1 - \xi_2)|} \, dy, \end{aligned}$$

which converges to zero as T_1, $T_2 \to \infty$, since the integral is finite. Moreover

$$
\begin{aligned}
|J_{22}| &\leqq (T_1T_2)^{1/2}C \int_{(\xi_1-\xi_2)-\eta<y<(\xi_1-\xi_2)+\eta} \\
&\qquad\qquad \cdot \frac{1}{T_1 y}(|p(y-\xi_1)| + |p(-\xi_1)|)\,dy \\
&\leqq (T_2/T_1)^{1/2}\frac{2C}{|\xi_1-\xi_2|}\int_{(\xi_1-\xi_2)-\eta}^{(\xi_1-\xi_2)+\eta}(|p(y-\xi_1)| + |p(-\xi_1)|)\,dy \\
&\leqq C\int_{(\xi_1-\xi_2)-\eta}^{(\xi_1-\xi_2)+\eta}(|p(y-\xi_1)| + |p(-\xi_1)|)\,dy.
\end{aligned} \tag{3.17}
$$

Hence $\limsup_{T_1,T_2\to\infty,T_1/T_2\to\mu}$ of (3.17) is small for η small, that is

$$\lim J_{22} = 0. \tag{3.18}$$

From (3.16), (3.18) we obtain

$$\lim J_2 = 0,$$

which together with (3.15) gives $\lim J = 0$.

4. Proof of the theorem. We now proceed to prove the theorem stated in Section 2.

We start with the computation of

$$
\begin{aligned}
ES(T_1)S(T_2) &= \frac{1}{T_1T_2}E\left|\int_{\infty}^{\infty}\varepsilon(t)M\left(\frac{t}{T_1}\right)e^{-i\xi t}\,dt\right|^2\cdot\left|\int_{\infty}^{\infty}\varepsilon(t)M\left(\frac{t}{T_2}\right)e^{-i\xi t}\,dt\right|^2 \\
&= \frac{1}{T_1T_2}E\int_{E_4}\varepsilon(t_1)\varepsilon(t_2)\varepsilon(t_3)\varepsilon t_4)e^{-i\xi(t_1-t_2+t_3-t_4)} \\
&\qquad\qquad \cdot M\left(\frac{t_1}{T_1}\right)\bar{M}\left(\frac{t_2}{T_1}\right)M\left(\frac{t_3}{T_2}\right)\bar{M}\left(\frac{t_4}{T_2}\right)dv_t \\
&= \frac{1}{T_1T_2}\int_{E_4}P(t_2-t_1,\, t_3-t_1,\, t_4-t_1)e^{-i\xi(t_1-t_2+t_3-t_4)} \\
&\qquad\qquad \cdot M\left(\frac{t_1}{T_1}\right)\bar{M}\left(\frac{t_2}{T_1}\right)M\left(\frac{t_3}{T_2}\right)\bar{M}\left(\frac{t_4}{T_2}\right)dv_t \\
&= \frac{1}{T_1T_2}\int_{E_4}Q(t_2-t_1,\, t_3-t_1,\, t_4-t_1)e^{-i\xi(t_1-t_2+t_3-t_4)} \\
&\qquad\qquad \cdot M\left(\frac{t_1}{T_1}\right)\bar{M}\left(\frac{t_2}{T_1}\right)M\left(\frac{t_3}{T_2}\right)\bar{M}\left(\frac{t_4}{T_2}\right)dv_t \\
&\quad + \frac{1}{T_1T_2}\int_{E_4}P_G(t_2-t_1,\, t_3-t_1,\, t_4-t_1)e^{-i\xi(t_1-t_2+t_3-t_4)} \\
&\qquad\qquad \cdot M\left(\frac{t_1}{T_1}\right)\bar{M}\left(\frac{t_2}{T_1}\right)M\left(\frac{t_3}{T_2}\right)\bar{M}\left(\frac{t_4}{T_2}\right)dv_t \\
&= S_1(T_1, T_2) + S_2(T_1, T_2).
\end{aligned}
$$

Inserting (2.6) in $S_1(T_1, T_2)$, we have

$$S_1(T_1, T_2) = (2\pi)^{-3} \frac{1}{T_1 T_2} \int_{E_4} M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) \cdot e^{-i\xi(t_1-t_2+t_3-t_4)}\, dv_t$$

$$\cdot \int_{E_3} q(x_1, x_2, x_3) \exp\,[i(t_2 - t_1)x_1 + i(t_3 - t_1)x_2 + i(t_4 - t_1)x_3]\, dv_x$$

$$(4.1) \qquad = (2\pi)^{-3} \frac{1}{T_1 T_2} \int_{E_3} q(x_1, x_2, x_3)\, dv_x \cdot \int_{E_4} M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right)$$

$$\cdot \exp\,[i\{-t_1(x_1 + x_2 + x_3 + \xi) + t_2(x_1 + \xi) + t_3(x_2 - \xi) + t_4(x_3 + \xi)\}]\, dv_t$$

$$= (2\pi)^{-3} T_1 T_2 \int_{E_3} q(x_1, x_2, x_3) K[-T_1(x_1 + x_2 + x_3 + \xi)] \cdot \bar{K}[-T_1(x_1 + \xi)] \cdot K[T_2(x_2 - \xi)] \bar{K}[-T_2(x_3 + \xi)]\, dv_x$$

$$= (2\pi)^{-3} T_1 T_2 \int_{E_3} q(x_1 - \xi, x_2 + \xi, x_3 - \xi) \cdot K[-T_1(x_1 + x_2 + x_3)] \cdot \bar{K}(-T_1 x) K(T_2 x_2) \bar{K}(-T_2 x_3)\, dv_x,$$

where we denote

$$(4.2) \qquad K(x) = \int_{\infty}^{\infty} M(\alpha) e^{ix\alpha}\, d\alpha.$$

Since $M(x)$ and $q(x_1, x_2, x_3)$ satisfy the condition of Lemma 1, we obtain that (4.1) multiplied by T_2 is convergent when $T_1/T_2 \to \mu (\mu \neq 0)$ Hence (4 1) converges to zero.

Next we shall consider $S_2(T_1, T_2)$ Inserting (2 3), we obtain

$$S_2(T_1, T_2) = \frac{1}{T_1 T_2} \int_{E_4} \{\rho(t_2 - t_1)\rho(t_3 - t_4) + \rho(t_3 - t_1)\rho(t_4 - t_2)$$

$$(4.4) \qquad + \rho(t_4 - t_1)\rho(t_2 - t_3)\} \bar{M}\left(\frac{t_1}{T_1}\right) M\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) \cdot e^{-i\xi(t_1-t_2+t_3-t_4)} \cdot dv_t$$

$$= U_1(T_1, T_2) + U_2(T_1, T_2) + U_3(T_1, T_2),$$

say, where

$$(4.5) \qquad U_1(T_1, T_2) = \frac{1}{T_1 T_2} \int_{E_4} \rho(t_2 - t_1)\rho(t_3 - t_4) \cdot M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) \cdot e^{-i\xi(t_1-t_2+t_3-t_4)}\, dv_t,$$

and U_2, U_3 are similar terms. By the assumptions of the theorem, we have $\rho(u) = \int_{-\infty}^{\infty} e^{iux}p(x)\,dx$, and, if we insert this into (4.5), we obtain

$$U_1(T_1, T_2) = \frac{1}{T_1T_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x)p(y)\,dx\,dy$$

$$\cdot \int_{E_4} M\left(\frac{t_1}{T_1}\right)e^{-it_1(x+\xi)}\bar{M}\left(\frac{t_2}{T_1}\right)e^{it_2(x+\xi)}M\left(\frac{t_3}{T_2}\right)e^{it_3(y-\xi)}\bar{M}\left(\frac{t_4}{T_2}\right)e^{-it_4(y-\xi)}\,dv_t$$

$$= T_1T_2\int_{-\infty}^{\infty} p(x)K[-T_1(x+\xi)]\bar{K}[-T_1(x+\xi)]\,dx$$

$$\cdot \int_{-\infty}^{\infty} p(y)K[T_2(y-\xi)]\bar{K}[T_2(y-\xi)]\,dy.$$

Since $\mathfrak{L}(t)$ is real, ρ is real too, and $p(x)$ is an even function, and hence by Lemma 2, we get

$$\lim_{T_1,T_2\to\infty} U_1(T_1, T_2) = (2\pi)^2C_1^2p(\xi)p(-\xi) = (2\pi C_1)^2p^2(\xi), \tag{4.6}$$

where

$$C_1 = \int_{-\infty}^{\infty}\bar{M}(\beta)M(\beta)\,d\beta = \int_{-\infty}^{\infty}|M(\beta)|^2\,d\beta. \tag{4.7}$$

Quite similarly

$$U_2(T_1, T_2) = T_1T_2\int_{-\infty}^{\infty} p(x)K[-T_1(x+\xi)]K[T_2(x-\xi)]\,dx$$

$$\cdot \int_{-\infty}^{\infty} p(y)\bar{K}[T_1(y-\xi)]\bar{K}[-T_2(y+\xi)]\,dy.$$

If $\xi = 0$, then, by (3.3),

$$U_2(T_1, T_2) \to (2\pi)^2\,|C_\mu^{(1)}|^2p^2(0), \qquad (T_1, T_2\to\infty,\ T_1/T_2\to\mu), \tag{4.8}$$

where

$$C_\mu^{(1)} = \mu^{1/2}\int_{-\infty}^{\infty} M(\beta)M(\mu\beta)\,d\beta. \tag{4.9}$$

If $\xi \neq 0$, then (3.4) shows

$$U_2(T_1, T_2)\to 0, \quad (T_1, T_2\to\infty,\ T_1/T_2\to\mu). \tag{4.10}$$

Finally we have

$$U_3(T_1, T_2) = T_1T_2\int_{-\infty}^{\infty} p(x)K[-T_1(x+\xi)]\bar{K}[-T_2(x+\xi)]\,dx$$

$$\cdot \int_{-\infty}^{\infty} p(y)\bar{K}[-T_1(y+\xi)]K[-T_2(y+\xi)]\,dy,$$

and

$$U_3(T_1, T_2)\to(2\pi)^2\,|C_\mu^{(2)}|^2p^2(\xi), \qquad \text{for every } \xi, \tag{4.11}$$

where

(4.12) [illegible]

Inserting (4.5), (4.8), (4.10) and (4.11) into (4.4), we get, if $\xi \neq 0$,

(4.13) $S(T_1, T_2)$ [illegible]

as $T_1, T_2 \to \infty$, $T_1/T_2 \to \alpha$ and if $\xi = 0$,

(4.14) $S(T_1, T_2)$ [illegible]

Hence we get

(4.15) $ES(T_1)S(T_2)$ [illegible] if $\xi \neq 0$,

(4.16) $ES(T_1)S(T_2)$ [illegible] if $\xi = 0$.

We also have

$$ES(T) = [illegible]$$

and by Lemma 2 this converges to [illegible]. Thus we find that

$$\operatorname{cov}\{S(T_1), S(T_2)\} = ES(T_1)S(T_2) - ES(T_1)\,ES(T_2)$$

converges to

[illegible] if $\xi \neq 0$,

and to

[illegible] if $\xi = 0$,

when T_1, T_2 increase indefinitely such as $T_1/T_2 \to \alpha$ ($\alpha > 0$).

Especially var $S(T)$ converges to

[illegible] if $\xi \neq 0$,

and to

$$((2\pi)^2(|C_1^{(0)}|^2 + |C_2^{(0)}|^2)p^2(0), \quad \text{if } \xi = 0.$$

Also, we easily find that

$$E\,|S(T_1) - S(T_2)|^2$$

converges to

$$2(2\pi)^2(|C_1^{(0)}|^2 - |C_2^{(0)}|^2)p^2(\xi), \quad \text{if } \xi \neq 0$$

and to

$$2(2\pi)^2(|C_1^{(0)}|^2 + |C_1^{(0)}|^2 - |C_2^{(0)}|^2 - |C_2^{(0)}|^2)p^2(0), \quad \text{if } \xi = 0$$

Hence the theorem is proved.

states, is a number h, called the *entropy* of the process, such that for large N it is practically certain that the sequence of states of length N which occurs is one whose probability is about 2^{-Nh}; more precisely, for any sequence $f \varepsilon F^{(N)}$ let

$$Q_N(f) = \text{Prob}\{(z_1, \cdots, z_N) = f\}.$$

Then

$$N^{-1} \log Q_N(z_1, \cdots, z_N) \to -h \text{ in } L_1 \text{ as } N \to \infty, \tag{1}$$

where the log above and throughout this paper has base 2. Breiman [1] has shown that convergence with probability 1 also occurs in (1). For the ergodic process $\{(d_k, r_k)\}$, the processes $\{x_k = \phi(d_k)\}$, $\{y_k = \psi(r_k)\}$, $\{(x_k, y_k)\}$ are of course also ergodic; we denote their entropies by $H(X)$, $H(Y)$, $H(X, Y)$ respectively.

For a fixed indecomposable channel, the upper bound H over all sources of the number $H(X) + H(Y) - H(X, Y)$ is called, following Shannon, the *capacity* of the channel. Shannon [5] and, subsequently, McMillan [4], Feinstein [2], Hincin [3], and Wolfowitz [6] have shown that, under various hypotheses on the channel, it is possible to transmit over the channel at any rate less than its capacity, but not at any rate greater than its capacity. For a channel as defined above, this means, as in [6], the following. For a given channel, to say that it is possible to transmit at rate G means that for every $\epsilon > 0$ there is an N_0 such that for any $N \geqq N_0$ there are $2^{NG} = J$ distinct sequences $u_1, \cdots, u_J$, where each $u_j \varepsilon I(A)^{(N)}$, and J disjoint subsets $E_1, \cdots, E_J$ of $I(B)^N$ such that

$$Q(r, u_j, E_j) > 1 - \epsilon \text{ for all } j \text{ and all } r \varepsilon I(R), \tag{2}$$

where for any $r \varepsilon I(R)$, $u = (u(1), \cdots, u(N)) \varepsilon I(A)^{(N)}$, $E \subset I(B)^{(N)}$

$$Q(r, u, E) = \sum C(u(1), r, r_1) \cdots C(u(N), r_{N-1}, r_N),$$

where the sum is over those sequences $(r_1, \cdots, r_N)$ for which

$$(\psi(r_1), \cdots, \psi(r_N)) \varepsilon E.$$

Thus $Q(r, u, E)$ is the probability that the output sequence from the channel is an element of E, when the channel is initially in state r and u is the input sequence.

For a given channel, denote by H^* the upper bound of the rates G at which it is possible to transmit. We shall show that, for indecomposable channels of the type considered here, $H^* = H$, that is, it is possible to transmit at any rate less than the channel capacity, but not at a rate greater than channel capacity. Shannon and McMillan seem to have regarded $H^* \leqq H$ as more or less obvious, and devoted most of their attention to showing, under certain hypotheses, that $H \leqq H^*$. The other writers have given some attention to the inequality $H^* \leqq H$. In particular, Wolfowitz [6] obtained $H^* \leqq H$ for channels of zero memory. Our result, that $H^* = H$ for indecomposable channels, extends those obtained previously.

3. A necessary and sufficient condition for indecomposability. To verify that the results to be proved in Sections 5 and 6 are valid for a given channel, we must show that the channel is indecomposable. The following criterion is helpful.

THEOREM 1. *A channel* $(C(1), \cdots, C(A))$ *is indecomposable if and only if every finite product* $C(a_1) \cdots C(a_k)$ *is an indecomposable Markov matrix,* $k = 1, 2, \cdots$, $a_i \in I(A)$.

PROOF. Suppose the channel is indecomposable and let $a_1, \cdots, a_k$ be any finite sequence of elements of $I(A)$. Consider the source with k states $1, \cdots, k$ with $M(i, i + 1) = 1$ for $i < k$, $M(k, 1) = 1$, and $\phi(i) = a_i$. Let

$$F = C(a_1) \cdots C(a_k)$$

and let $r_1, r_2 \in I(R)$. To show that F is indecomposable it is sufficient to find integers T_1, T_2 and a state $r_3 \in I(R)$ such that $F^{T_1}(r_1, r_3) > 0$ and $F^{T_2}(r_2, r_3) > 0$, that is, such that r_3 is reachable from either r_1 or r_2 under transition matrix F. Since the source-channel matrix L is indecomposable, the two states (k, r_1), (k, r_2) have a common possible successor (i, r) which itself has a possible successor of the form (k, r_3). Thus (k, r_3) is a possible successor of either (k, r_1) or (k, r_2). Since the source has period k, the times after which (k, r_3) can be reached from (k, r_1) or (k, r_2) are multiples of k, that is, there are integers T_1, T_2 such that $L^{T_i k}((k, r_i), (k, r_3)) > 0$ for $i = 1, 2$. But $L^{Tk}((k, r), (k, s)) = F^T(r, s)$. Consequently $F^{T_i}(r_i, r_3) > 0$ for $i = 1, 2$ and F is indecomposable.

Now suppose that every finite product $C(a_1) \cdots C(a_k)$ is indecomposable, and let (M, ϕ) be any source. Let (d, r), (e, s) be any two source-channel states; we must find a common possible successor (f, t). Since M is indecomposable, d and e have a common possible successor f which is recurrent. There are then numbers r', s', such that (f, r') is a successor of (d, r) and (f, s') is a successor of (e, s), so that any common successor of (f, r') and (f, s') is also a common successor of (d, r) and (e, s). Thus we may suppose $d = e = f$, and must find a common successor (f, t) of (f, r'), (f, s'), where f is recurrent. Let $f_0 = f, f_1, \cdots, f_{k-1}, f_k = f$ be a possible path from f to itself, and let $F = C(\phi(f_1)) \cdots C(\phi(f_k))$. We assert that if t is a possible successor of r' with respect to F, then (f, t) is a possible successor of (f, r') in the source-channel matrix L. For $L^{Tk}((f, r'), (f, t)) \geqq [M(f_0, f_1) \cdots M(f_{k-1}, f_k)]^T F^T(r', t)$, and since the first factor on the right is positive, the left side is positive whenever $F^T(r', t)$ is. But since F is recurrent, r' and s' have a common possible successor t with respect to F, so that (f, t) is a common possible successor of (f, r'), (f, s) in L, completing the proof.

We shall say that a channel has memory m if every product $C(a_0) \cdots C(a_m)$ has identical rows. Thus a channel has memory m if and only if the conditional distribution of the present state of the channel, given the present input a_m, the m previous inputs $a_0, \cdots, a_{m-1}$ and the state r of the channel just prior to input a_0, is independent of r for every $a_0, \cdots, a_m$. A channel is said to have finite memory if for some m it has memory m. Every channel with finite memory is clearly indecomposable, for if $F = C(a_1) \cdots C(a_k)$, some power of F has

number h, called the *entropy* of the process, such that for large N it
ly certain that the sequence of states of length N which occurs is one
ability is about 2^{-Nh}; more precisely, for any sequence $f \varepsilon F^{(N)}$ let

$$Q_N(f) = \text{Prob}\,\{(z_1, \cdots, z_N) = f\}.$$

$$N^{-1} \log Q_N(z_1, \cdots, z_N) \to -h \text{ in } L_1 \text{ as } N \to \infty,$$

log above and throughout this paper has base 2. Breiman [1] has
convergence with probability 1 also occurs in (1). For the ergodic
$_k, r_k)\}$, the processes $\{x_k = \phi(d_k)\}$, $\{y_k = \psi(r_k)\}$, $\{(x_k, y_k)\}$ are of
ergodic; we denote their entropies by $H(X)$, $H(Y)$, $H(X, Y)$ re-

d indecomposable channel, the upper bound H over all sources of the
$X) + H(Y) - H(X, Y)$ is called, following Shannon, the *capacity*
nel. Shannon [5] and, subsequently, McMillan [4], Feinstein [2],
and Wolfowitz [6] have shown that, under various hypotheses on the
is possible to transmit over the channel at any rate less than its
it not at any rate greater than its capacity. For a channel as defined
means, as in [6], the following. For a given channel, to say that it is
ransmit at rate G means that for every $\epsilon > 0$ there is an N_0 such that
$\geq N_0$ there are $2^{NG} = J$ distinct sequences $u_1, \cdots, u_J$, where each
and J disjoint subsets $E_1, \cdots, E_J$ of $I(B)^N$ such that

$$Q(r, u_j, E_j) > 1 - \epsilon \text{ for all } j \text{ and all } r \varepsilon I(R),$$

y $r \varepsilon I(R)$, $u = (u(1), \cdots, u(N)) \varepsilon I(A)^{(N)}$, $E \subset I(B)^{(N)}$

$$Q(r, u, E) = \sum C(u(1), r, r_1) \cdots C(u(N), r_{N-1}, r_N),$$

m is over those sequences $(r_1, \cdots, r_N)$ for which

$$(\psi(r_1), \cdots, \psi(r_N)) \varepsilon E.$$

$E)$ is the probability that the output sequence from the channel is an
, when the channel is initially in state r and u is the input sequence.
n channel, denote by H^* the upper bound of the rates G at which it
transmit. We shall show that, for indecomposable channels of the
red here, $H^* = H$, that is, it is possible to transmit at any rate less
nnel capacity, but not at a rate greater than channel capacity.
McMillan seem to have regarded $H^* \leqq H$ as more or less obvious,
most of their attention to showing, under certain hypotheses, that
other writers have given some attention to the inequality $H^* \leqq H$.
r, Wolfowitz [6] obtained $H^* \leqq H$ for channels of zero memory.
that $H^* = H$ for indecomposable channels, extends those ob-
usly.

3. A necessary and sufficient condition for indecomposability. To verify that the results to be proved in Sections 5 and 6 are valid for a given channel, we must show that the channel is indecomposable. The following criterion is helpful.

THEOREM 1. *A channel* $(C(1), \cdots, C(A))$ *is indecomposable if and only if every finite product* $C(a_1) \cdots C(a_k)$ *is an indecomposable Markov matrix,* $k = 1, 2, \cdots$, $a_i \in I(A)$.

PROOF. Suppose the channel is indecomposable and let $a_1, \cdots, a_k$ be any finite sequence of elements of $I(A)$. Consider the source with k states $1, \cdots, k$ with $M(i, i + 1) = 1$ for $i < k$, $M(k, 1) = 1$, and $\phi(i) = a_i$. Let

$$F = C(a_1) \cdots C(a_k)$$

and let $r_1, r_2 \in I(R)$. To show that F is indecomposable it is sufficient to find integers T_1, T_2 and a state $r_3 \in I(R)$ such that $F^{T_1}(r_1, r_3) > 0$ and $F^{T_2}(r_2, r_3) > 0$, that is, such that r_3 is reachable from either r_1 or r_2 under transition matrix F. Since the source-channel matrix L is indecomposable, the two states (k, r_1), (k, r_2) have a common possible successor (i, r) which itself has a possible successor of the form (k, r_3). Thus (k, r_3) is a possible successor of either (k, r_1) or (k, r_2). Since the source has period k, the times after which (k, r_3) can be reached from (k, r_1) or (k, r_2) are multiples of k, that is, there are integers T_1, T_2 such that $L^{T_i k}((k, r_i), (k, r_3)) > 0$ for $i = 1, 2$. But $L^{Tk}((k, r), (k, s)) = F^T(r, s)$. Consequently $F^{T_i}(r_i, r_3) > 0$ for $i = 1, 2$ and F is indecomposable.

Now suppose that every finite product $C(a_1) \cdots C(a_k)$ is indecomposable, and let (M, ϕ) be any source. Let (d, r), (e, s) be any two source-channel states, we must find a common possible successor (f, t). Since M is indecomposable, d and e have a common possible successor f which is recurrent. There are then numbers r', s', such that (f, r') is a successor of (d, r) and (f, s') is a successor of (e, s), so that any common successor of (f, r') and (f, s') is also a common successor of (d, r) and (e, s). Thus we may suppose $d = e = f$, and must find a common successor (f, t) of (f, r'), (f, s'), where f is recurrent. Let $f_0 = f, f_1, \cdots, f_{k-1}, f_k = f$ be a possible path from f to itself, and let $F = C(\phi(f_1)) \cdots C(\phi(f_k))$. We assert that if t is a possible successor of r' with respect to F, then (f, t) is a possible successor of (f, r') in the source-channel matrix L. For $L^{Tk}((f, r'), (f, t)) \geqq [M(f_0, f_1) \cdots M(f_{k-1}, f_k)]^T F^T(r', t)$, and since the first factor on the right is positive, the left side is positive whenever $F^T(r', t)$ is. But since F is recurrent, r' and s' have a common possible successor t with respect to F, so that (f, t) is a common possible successor of (f, r'), (f, s) in L, completing the proof.

We shall say that a channel has memory m if every product $C(a_0) \cdots C(a_m)$ has identical rows. Thus a channel has memory m if and only if the conditional distribution of the present state of the channel, given the present input a_m, the m previous inputs $a_0, \cdots, a_{m-1}$ and the state r of the channel just prior to input a_0, is independent of r for every $a_0, \cdots, a_m$. A channel is said to have finite memory if for some m it has memory m. Every channel with finite memory is clearly indecomposable, for if $F = C(a_1) \cdots C(a_k)$, some power of F has

nd rows so that P is indecomposable. From Theorem 1, the channel is then mposable. That this includes, as a special case, the finite memory channels ned by Feinstein [2] and Wolfowitz [6] can be seen from the following rations: let the inputs to a channel be denoted by $\cdots, X_{-1}, X_0, X_1, \cdots$ e outputs by $\cdots, Y_{-1}, Y_0, Y_1, \cdots$ and let the probability structure of annel be defined, following McMillan [4], by specifying the conditional bilities of the various output messages, given the input signals. That is, we on the conditional probabilities $p(Y_n, \cdots, Y_k \mid X_n, X_{n-1}, \cdots)$ where now assuming that the channel is nonanticipatory and stationary. We , in addition, that there is an integer m such that

$$p(Y_n \mid X_n, Y_{n-1}, X_{n-1}, Y_{n-2}, X_{n-2}, \cdots) = p(Y_n \mid X_n, Y_{n-1}, X_{n-1}, \cdots, Y_{n-m}, X_{n-m}).$$

we consider the finite state channel whose states consist of m-tuples of one member of the pairs being from the input alphabet and the other from tput alphabet, then the above assumption implies that this finite state l is finitary in the sense described above, that is, it has the required y property. If we add the additional restriction that there is an integer that if two output messages m long, say y_1, y_2, are separated by a dis- M, that

$$p(y_1, y_2 \mid \cdots X_1, X_0, X_{-1}, \cdots) = p(y_1 \mid \cdots, X_1, X_0, X_{-1}, \cdots)p(Y_2 \mid \cdots X_1, X_0, X_{-1})$$

is finite state channel has finite memory M.

modification of McMillan's theorem. In proving our main result, we ed the following extension of a special case of McMillan's theorem.

orem 2. *Let $d_1, d_2, \cdots$ be a Markov process with finite indecomposable on matrix M, say $D \times D$, let ϕ be a function from $I(D)$ to $I(A)$, and let (d_n). For any sequence $a \in I(A)^{(N)}$ let $p(a) = P\{(y_1, \cdots, y_N) = a\}$, and let $(y_1, \cdots, y_N)$. There is a constant h, depending only on M and ϕ, such that*

$$N^{-1} \log z_N \to -h$$

nd with probability 1 as $N \to \infty$.

of. If the distribution of d_1 is the (unique) stationary distribution for M, } process is ergodic, and the theorems of McMillan [4] and Breiman [1]), with h as the entropy of the process.

any $d \in I(D)$ and any event B, write $P_d(B)$ for $P(B \mid d_1 = d)$. Let $\lambda = \cdots, \lambda_D)$ be the stationary distribution for M, and let $Q(B) = \sum \lambda_d P_d(B)$. eorems of McMillan and Breiman assert that

$$\frac{\log \sum_d \lambda_d z_{dN}}{N} \to -h \text{ a.s. and } L_1(Q),$$

where $p_d(s) = P_d\{(y_1, \cdots, y_N) = s\}$ and $z_{dN} = p_d(y_1, \cdots, y_N)$. For any d for which $\lambda_d > 0$, we have

$$\lambda_d z_{dN} = (\sum_e \lambda_e z_{eN}) Q(d_1 = d \mid y_1, \cdots, y_N).$$

Taking logs, dividing by N, letting $N \to \infty$ and using (4) and the fact that $Q(d_1 = d \mid y_1, \cdots, y_N)$ converges a.e. (Q) to a limit which is positive a.e. (P_d) yields

(5) $$\frac{\log z_{dN}}{N} \to -h \text{ a.e } P_d \text{ for } \lambda_d > 0.$$

Now let d be a state for which $\lambda_d = 0$, let e be any state for which $\lambda_e > 0$, let k be any integer $\leqq N$ and let G denote the event $\{d_k = e\}$. We have

(6) $$z_{dN}P_d(G \mid y_1, \cdots, y_N) = z_{dk}P_d(G \mid y_1, \cdots, y_k)p_e(y_k, \cdots, y_N)$$

Since the P_d conditional distribution of $y_k, y_{k+1}, \cdots$, given that G occurs, is the same as the unconditional P_e distribution of $y_1, y_2, \cdots$, we conclude from (5) that on G, a.e. P_d, $N^{-1} \log p_e(y_k, \cdots, y_N) \to -h$. Also, on G, a.e. P_d, $P_d(G \mid y_1, \cdots, y_N)$ has a positive limit as $N \to \infty$ and $z_{dk}P_d(G \mid y_1, \cdots, y_k)$ is positive. Taking logs in (6), dividing by N, letting $N \to \infty$ yields

(7) $$N^{-1} \log z_{dN} \to -h \text{ a.e. } P_d \text{ on } G.$$

Since the union of the sets G obtained by varying k and e has P_d measure 1, we conclude

(8) $$N^{-1} \log z_{dN} \to -h \text{ a.e. } P_d \text{ for all } d.$$

Next let $\mu = (\mu_1, \cdots, \mu_D)$ be any initial distribution and let $P = \sum \mu_d P_d$ For any d for which $\mu_d > 0$, we have

(9) $$\mu_d z_{dN} = (\sum_d \mu_d z_{dN}) P(d_1 = d \mid y_1, \cdots, y_N).$$

Taking logs, dividing by N, letting $N \to \infty$ and using (8) yields

(10) $$N^{-1} \log (\sum \mu_d z_{dN}) \to -h \text{ a.e. } P_d,$$

from which we obtain

(11) $$N^{-1} \log (\sum \mu_d z_{dN}) \to -h \text{ a.e. } P.$$

Thus the probability 1 convergence in (2) is established. Finally, to obtain L_1 convergence we note, following McMillan, that the sequence $\{N^{-1} \log z_N\}$ is uniformly integrable. We have

(12) $$J(N,k) = \int_{B_k} \mid N^{-1} \log z_N \mid dP = N^{-1} \sum p(s) \mid \log p(s) \mid \leqq (k+1)2^{-kN}A^N,$$

where B_k is the event $\{k \leqq |N^{-1} \log z_N| < k + 1\}$ and the sum is extended over those $s \,\varepsilon\, I(A)^{(N)}$ for which $k \leqq |N^{-1} \log p(s)| < k + 1$. Choose k_1 so that $2^{-k_1} A < 1$. For $k \geqq k_1$ we have

$$J(N, k) \leqq (k + 1)2^{-k}2^{k_1}. \tag{13}$$

Thus $\sum_{k_0}^{\infty} J(N, k)$ goes to zero as $k_0 \to \infty$ uniformly in N, and uniform integrability is established, completing the proof.

5. The direct half of Shannon's theorem (possibility of transmission at every rate less than capacity). We shall need the following lemma.

LEMMA. *Let p be a probability distribution on a finite product space $X \times Y$. Write $a(x) = \sum_y p(x, y)$, $b(y) = \sum_x p(x, y)$, $p(y \mid x) = p(x, y) / a(x)$. For any numbers δ, λ such that $0 < \delta \leqq \lambda < 1$, let*

$$A = \{y : b(y) > \delta\}, \qquad B = \{(x, y) : p(y \mid x) < \lambda\}.$$

For any integer M there are M points $x_1, \cdots, x_M \,\varepsilon\, X$ and M disjoint subsets $E_1, \cdots, E_M$ of Y such that

$$\sum_{y \notin E_i} p(y \mid x_i) \leqq 4M(\delta/\lambda) + 2 \sum_{y \varepsilon A} b(y) + 2 \sum_{(x,y) \varepsilon B} p(x, y) \tag{14}$$

for $i = 1, \cdots, M$.

PROOF. Let $X_1, \cdots, X_{2M}$ be independent random variables with distribution $a(x)$. For each $i \,\varepsilon\, I(2M)$, $y \,\varepsilon\, Y$, we define the random variable $Z(i, y) = 1$ if $p(y \mid X_i) \leqq \max_{j \neq i} p(y \mid X_j)$, $Z(i, y) = 0$ otherwise, and define

$$f_i = \sum_y p(y \mid X_i) Z(i, y).$$

Then

$$\begin{aligned} Ef_i &= \sum_x a(x) E(f_i \mid X_i = x) = \sum_{x,y} p(x, y) E(Z(i, y) \mid X_i = x) \\ &\leqq \sum_{y \varepsilon A} b(y) + \sum_{(x,y) \varepsilon B} p(x, y) + \sum{}^{*} p(x, y) E(Z(i, y) \mid X_i = x), \end{aligned} \tag{15}$$

where $\sum^{*}$ indicates summation over pairs (x, y) for which $b(y) \leqq \delta$ and $p(y \mid x) \geqq \lambda$. Now $E(Z(i, y) \mid X_i = x) = 1 - (1 - u(x, y))^{2M-1}$, where

$$u(x,y) = \sum_{v \ni p(y|v) \geqq p(y|x)} a(v).$$

For pairs (x, y) in $\sum^{*}$,

$$\delta \geqq b(y) = \sum_v a(v) p(y \mid v) \geqq \lambda \sum_{p(y|v) \geqq \lambda} a(v) \geqq \lambda u(x, y),$$

so that

$$E(Z(i, y) \mid X_i = x) \leqq 1 - (1 - (\delta/\lambda))^{2M-1} \leqq 2M\delta/\lambda.$$

Using this inequality in (15) yields

$$Ef_i \leqq \sum_{y \varepsilon A} b(y) + \sum_{x,y \varepsilon B} p(x,y) + 2M\delta/\lambda = \alpha. \tag{16}$$

It follows that $E(\sum_1^{2M} f_i / 2M) \leqq \alpha$. Thus there are values of $X_1, \cdots, X_{2M}$, say $x_1^*, \cdots, x_{2M}^*$, for which $\sum_1^{2M} f_i^* / 2M \leqq \alpha$, where $f_i^* = f_i(x_1^*, \cdots, x_{2M}^*)$. Since all f_i^* are $\geqq 0$, at least M of them, say $f_{i_1}^*, \cdots, f_{i_M}^*$, are $\leqq 2\alpha$. Then $2\alpha \geqq \sum' p(y \mid x_{i_j}^*)$ where the sum is over y for which

$$p(y \mid x_{i_j}^*) \leqq \max_{i \neq i_j} p(y \mid x_i^*).$$

Denoting $x_{i_j}^*$ by x_j and the set of y for which

$$p(y \mid x_{i_j}^*) > \max_{i \neq i_j} p(y \mid x_i^*)$$

by E_j yields (14), and the lemma is proved

THEOREM 3. *For any indecomposable channel, $H^* \geqq H$, that is, it is possible to transmit at any rate less than the capacity of the channel.*

PROOF. Let (M, ϕ) be any source and let $\{(d_n, r_n), n = 0, 1, 2, \cdots\}$ be a Markov process whose transition matrix is the source-channel matrix and with d_0, r_0 having a uniform distribution on the DR states. Let $x_n = \phi(d_n)$, $y_n = \psi(r_n)$. For any $s \varepsilon I(A)^{(N)}$, $t \varepsilon I(B)^{(N)}$, $r \varepsilon I(R)$, write

$$a(s) = P((x_1, \cdots, x_N) = s), \qquad b(t) = P((y_1, \cdots, y_N) = t),$$

$$Q(r, s, t) = P((x_1, \cdots, x_N) = s, \quad (y_1, \cdots, y_N) = t, \quad r_0 = r)R/a(s),$$

$$p(s, t) = P((x_1, \cdots, x_N) = s, \quad (y_1, \cdots, y_N) = t) = a(s) \textstyle\sum_r Q(r, s, t)/R.$$

According to Theorem 2, as $N \to \infty$

$$N^{-1} \log a(x_1, \cdots, x_N) \to -H(X)$$

$$N^{-1} \log b(y_1, \cdots, y_N) \to -H(Y)$$

$$N^{-1} \log p(x_1, \cdots, x_N, y_1, \cdots, y_N) \to -H(X, Y).$$

Given $\epsilon > 0$, choose N so large that, with probability $\geqq 1 - \epsilon$,

$$\frac{\log p(x_1, \cdots, x_N, y_1, \cdots, x_N) - \log a(x_1, \cdots, x_N)}{N} \geqq H(X) - H(X,Y) - \epsilon$$

and

$$\frac{\log b(y_1, \cdots, y_N)}{N} \leqq -H(Y) + \epsilon.$$

We apply the lemma to the product space $U \times V$, where $U = I(A)^{(N)}$, $V = I(B)^{(N)}$, with $p(u, v)$ as defined above and $\delta = 2^{-N(H(Y)-\epsilon)}$, $\lambda = 2^{-N(H(X,Y)-H(X)-\epsilon)}$, and conclude the existence of $M = 2^{NG}$, say, points $u_1, \cdots, u_M \varepsilon X$ and M disjoint subsets $E_1, \cdots, E_M$ of V such that

$$\sum_{u \notin E_i} p(u \mid v_i) \leqq 4 \cdot 2^{-N[H(Y)+H(X)-H(X,Y)-G-2\epsilon]} + 8\epsilon.$$

Thus for any $G < H(X) + H(Y) - H(X, Y)$ we can, for any $\beta > 0$, by first choosing ϵ sufficiently small (less than min $(\beta/9, (H(X) + H(Y) - H(X, Y) - G)/2)$ and then choosing N sufficiently large, find $M = 2^{NG}$ X-sequences $u_1, \cdots, u_M$ of length N and M disjoint subsets $E_1, \cdots, E_M$ of $I(B)^{(N)}$ such that

$$\sum_{v \varepsilon E_i} p(v \mid u_i) > 1 - \beta. \tag{17}$$

This does not quite prove that it is possible to transmit at rate G as defined above, since (2) requires that

$$\sum_{v \varepsilon E_i} Q(r, u_i, v) > 1 - \epsilon \quad \text{for all} \quad r \varepsilon R,$$

that is, that for each initial state of the channel, each of the M messages can be correctly recovered, with large probability. This is an immediate consequence of (17), however, since (17) yields

$$R^{-1} \sum_r \left(\sum_{v \varepsilon E_i} Q(r, u_i, v)\right) > 1 - \beta,$$

so that, since $Q(r, u_i, E_i) \leqq 1$ for all r, i,

$$\sum_{v \varepsilon E_i} Q(r, u_i, v) > 1 - R\beta$$

for each r. Since β can be made arbitrarily small and R is a fixed number, the number of states of the channel, the proof is complete.

6. The converse half of Shannon's theorem (impossibility of transmission at a rate greater than capacity).

THEOREM 4. *For any indecomposable channel, $H^* \leqq H$, that is, it is not possible to transmit at a rate greater than the capacity of the channel.*

PROOF. Suppose that it is possible to transmit over a given channel at rate G, let ϵ be given, $0 < \epsilon < \frac{1}{2}$ and let $N, u_1, \cdots, u_J, J = 2^{NG}, E_1, \cdots, E_J$ denote the quantities whose existence is implied by the possibility of transmission at rate G. We may suppose that $\bigcup E_j = I(B)^{(N)}$, since if (2) is satisfied for E_j it is also satisfied if E_j is replaced by a superset. We must exhibit a source (M, ϕ) for which $H(X) + H(Y) - H(X, Y)$ is nearly G. Our source produces inputs in blocks of N by selecting one of the u_j at random, successive choices being independent. The entropy $H(X)$ will then be precisely G. Since observing a long y sequence nearly identifies the corresponding x sequence, the conditional entropy $H(X, Y) - H(Y)$ is small, so that $H(X) + H(Y) - H(X, Y)$ is nearly G.

More precisely, the input source will have NJ states (n, j), with $M((n, j), (n + 1, j)) = 1$ for $n < N$, $M((N, j), (1, i)) = 1/J$ for $i \varepsilon I(J)$. We define $\phi(n, j) = u_{jn}$, the nth symbol in the sequence u_j. Let (d_k, r_k) be a Markov process whose transition matrix is the source-channel matrix, and whose initial distribution is such that $d_1 = (1, i)$ with probability $1/J$, $i \varepsilon I(J)$ and write $x_k = \phi(d_k)$, $y_k = \psi(r_k)$. Then every x sequence of length NT which is possible has probability $J^{-T} = 2^{-NTG}$ (since $\epsilon < \frac{1}{2}$, $u_i \neq u_j$ for $i \neq j$). From Theorem 2, $H(X) = G$.

To estimate $H(X, Y) - H(Y)$, we recall some results of Shannon [5]. If x is any random variable assuming T distinct values with probabilities $p_1, \cdots, p_T$, the number $-\sum p_t \log p_t$ is called the *entropy* of x and will be denoted by $h(x)$. Always $h(x) \leqq \log T$. If (x, y) are two random variables, each with a finite set of values, the number $h(x, y) - h(y)$ is called the conditional entropy of x given y and is denoted by $h(x \mid y)$. It equals the expected value of the entropy of the conditional distribution of x given y. For any function ϕ defined on the range of y, $h(\phi(y)) \leqq h(y)$ and $h(x \mid \phi(y)) \geqq h(x \mid y)$.

Notice that, in the notation of Theorem 2, $E \log z_N = -h(y_1, \cdots, y_N)$, so that the L_1 convergence in (2) implies that $h(y_1, \cdots, y_N)/N \to H$ as $N \to \infty$. Thus, in our present notation,

$$h(x_1, \cdots, x_{NT} \mid y_1, \cdots, y_{NT})/NT \to H(X, Y) - H(Y)$$

as $T \to \infty$. We have

$$h(x_1, \cdots, x_{NT} \mid y_1, \cdots, y_{NT}) \leqq \sum_{t=0}^{T-1} h(x_{Nt+1}, \quad , x_{Nt+N} \mid y_{Nt+1}, \cdot \quad \cdot, y_{Nt+N})$$

$$\leqq \sum_{t=1}^{T-1} h(a_t \mid b_t),$$

where $a_t = (x_{Nt+1}, \cdots, x_{Nt+N})$ and $b_t = u_j$ if $(y_{Nt+1}, \cdots, y_{Nt+N}) \varepsilon E_j$ (we may suppose that $\cup E_j = I(B)^{(N)}$). We estimate $h(a_t \mid b_t)$ by the following lemma.

Lemma. *For any distribution α on a product space $U \times U$ of pairs (a, b) such that $\sum_j \alpha(a, a) \geqq 1 - \epsilon > \frac{1}{2}$ we have*

$$h(a \mid b) \leqq -g(\epsilon) + \epsilon \log (J - 1),$$

where $g(t) = t \log t + (1 - t) \log (1 - t)$, $0 \leqq t \leqq 1$, and J is the number of elements of U.

Proof of the Lemma. Let $\beta(b) = \sum_a \alpha(a, b)$. Then

$$-h(a \mid b) = \sum_b \beta(b) \sum_a \frac{\alpha(a, b)}{\beta(b)} \log \frac{\alpha(a, b)}{\beta(b)}.$$

Now

$$\sum_j \frac{\alpha(a, b)}{\beta(b)} \log \frac{\alpha(a, b)}{\beta(b)}$$

$$= \frac{\alpha(b, b)}{\beta(b)} \log \frac{\alpha(b, b)}{\beta(b)} + \frac{\beta(b) - \alpha(b, b)}{\beta(b)}$$

$$\cdot \sum_{a \neq b} \frac{\alpha(a, b)}{\beta(b) - \alpha(b, b)} \log \frac{\alpha(a, b)}{\beta(b) - \alpha(b, b)} + \frac{\beta(b) - \alpha(b, b)}{\beta(b)} \log \frac{\beta(b) - \alpha(b, b)}{\beta(b)}$$

$$= g\left(\frac{\alpha(b, b)}{\beta(b)}\right) - \frac{\beta(b) - \alpha(b, b)}{\beta(b)} \log (J - 1).$$

Consequently,

$$-h(a \mid b) \geqq \sum_b \beta(b) g\left(\frac{\alpha(b, b)}{\beta(b)}\right) - \epsilon \log (J - 1).$$

Since $g(t)$ is convex and $\sum_b \beta(b) = 1$,

$$\sum_b \beta(b) g\left[\frac{\alpha(b, b)}{\beta(b)}\right] \geqq g[\sum_b \alpha(b, b)] \geqq g(1 - \epsilon) = g(\epsilon).$$

The hypotheses of the lemma are satisfied for (a_t, b_t), so that

$$h(a_t \mid b_t) \leqq -g(\epsilon) + \epsilon \log J = -g(\epsilon) + \epsilon NG.$$

Thus

$$h(x_1, \cdots, x_{NT} \mid y_1, \cdots, y_{NT}) \leqq T(-g(\epsilon) + \epsilon NG).$$

Dividing by NT and letting $T \to \infty$ yields

$$H(X, Y) - H(Y) \leqq -\frac{g(\epsilon)}{N} + \epsilon G.$$

Thus, assuming that transmission at rate G is possible we have for every $\epsilon > 0$ and arbitrarily large N, exhibited a source for which

$$H(X) + H(Y) - H(X, Y) \geqq G(1 - \epsilon) + g(\epsilon)/N.$$

It follows that $H \geqq H^*$ and the proof is complete.

7. Another form of Shannon's Theorem. Let $\{w_n, n = 1, 2, \ldots\}$ be any stationary ergodic process whose variables have a finite set of values, say $I(W)$, and consider a given indecomposable channel as defined above. Shannon enquires whether the channel is adequate for transmitting the information produced by the source, with large probability of correct reception. To say that the channel is adequate means that, for every $\epsilon > 0$, there is an integer N_0 such that for any $N \geqq N_0$ there are (1) a function f (the encoder) from $I(W)^{(N)}$ to $I(A)^{(N)}$ and (2) a function g (the decoder) from $I(B)^{(N)}$ to $I(W)^N$ such that, for every initial state r of the channel,

$$\pi_r\{\alpha = \beta\} > 1 - \epsilon,$$

where α and β are random variables (the first N symbols produced by the source and the decoded estimate for these symbols respectively) whose joint distribution π_r is defined by

$$\pi_r\{\alpha = v, \beta = v'\} = \text{Prob}\ \{(w_1, \cdots, w_N) = v\} \sum_{g(\delta)=v'} Q(r, f(v), \delta)$$

where $Q(r, u, \delta)$, as defined earlier, is the probability that the channel, when initially in state r, on receiving an input u, will produce output δ. The form in which Shannon describes his result is the following.

THEOREM 5. *An indecomposable channel of capacity H is adequate for the stationary ergodic source $\{w_n\}$ if the entropy h of $\{w_n\}$ is less than H, and not if $h > H$.*

The idea of the proof of this result, based on McMillan's theorem and Theorems 3 and 4 above, is extremely simple. According to McMillan's theorem, the source w_n is very likely to produce one of about 2^{hN} sequences of length N, each of which has probability about 2^{-hN}. Accordingly, to have a large probability of transmitting the actual sequence accurately, it is necessary and sufficient that the channel be able to distinguish among about 2^{hN} different input sequences of length N which, by Theorem 3, it is if $h < H$ and is not if $h > H$. The proof below simply makes this idea precise.

PROOF. From (1), for any $\epsilon > 0$ there is an N_1 such that for any $N \geqq N_1$ there is a set $F \subset I(W)^{(N)}$ with not more than $2^{(h+\epsilon)N}$ elements such that

$$\text{Prob } \{(w_1, \cdots, w_N) \; \varepsilon \; F\} > 1 - \epsilon.$$

From Theorem 3 there is an $N_0 \geqq N_1$ such that for any $N \geqq N_0$ there are $2^{(H-\epsilon)N} = J$ distinct sequences $u_1, \cdots, u_J$ in $I(A)^{(N)}$ and J disjoint subsets $E_1, \cdots, E_J$ of $I(B)^{(N)}$ such that

$$Q(r, u_j, E_j) > 1 - \epsilon \text{ for all } j \text{ and } r$$

If $H - \epsilon \geqq h + \epsilon$ there are at most J elements in F, so that there is a function f from $I(W)^{(N)}$ to $I(A)^N$ such that f maps distinct elements of F onto distinct u_j. With this f and with g chosen so that $g(\delta) \; \varepsilon \; F$, $f[g(\delta)] = u_j$ for all $\delta \; \varepsilon \; E_j$, we have

$$\pi_r\{\alpha = \beta\} > (1 - \epsilon)^2,$$

since the probability that $\alpha \; \varepsilon \; F$ is greater than $1 - \epsilon$ and the conditional probability, given that $\alpha = \alpha_0 \; \varepsilon \; F$, that $\beta = \alpha_0$ is at least $Q(r, u_j, E_j) > 1 - \epsilon$, where $f(\alpha_0) = u_j$. Thus if $h < H$, the channel is adequate.

Conversely, suppose the channel is adequate. From (1), for any $\epsilon > 0$ there is an N_2 such that for any $N \geqq N_2$ there is a set $F_1 \subset I(W)^{(N)}$ such that

$$\text{Prob } \{(w_1, \cdots, w_N) \; \varepsilon \; F_1\} > 1 - \epsilon$$

and Prob $(w_1, \cdots, w_N) = \alpha_0 < 2^{-(h-\epsilon)N}$ for all $\alpha_0 \; \varepsilon \; F_1$. Also, there is an $N_3 \geqq N_2$ such that for every $N \geqq N_3$ there are functions f, g satisfying the definition of adequacy. Since $\pi_r\{\alpha = \beta\} > 1 - \epsilon$, there is a subset F_2 of $I(W)^{(N)}$ such that $\pi_r\{\alpha \; \varepsilon \; F_2\} > 1 - \sqrt{\epsilon}$ the conditional probability

$$\pi_r\{\alpha = \beta \mid \alpha = \alpha_0\} > 1 - \sqrt{\epsilon} \text{ for } \alpha_0 \; \varepsilon \; F_2 .$$

Then $\pi_r\{\alpha \; \varepsilon \; F_1 \cap F_2\} > 1 - \epsilon - \sqrt{\epsilon}$, so that $F_1 \cap F_2$, and hence F_2 has at least $2^{(h-\epsilon)N}(1 - \epsilon - \sqrt{\epsilon}) = J_1$ elements. For $\alpha_0 \; \varepsilon \; F_2$, define $E(\alpha_0)$ as the set of all $\delta \; \varepsilon \; I(B)^{(N)}$ such that $g(\delta) = \alpha_0$. The assertion $\pi_r\{\alpha = \beta \mid \alpha = \alpha_0\} > 1 - \sqrt{\epsilon}$ is equivalent to

$$Q(r, f(\alpha_0), E(\alpha_0) > 1 - \sqrt{\epsilon}. \tag{17}$$

Note that, since the sets $E(\alpha_0)$ are disjoint, so are the elements of (α_0), provided $\epsilon < .707$, which we may assume. In summary, for every $\epsilon > 0$ we have found an

N_3 such that for any $N \geqq N_3$ there are at least $J_1(N, \epsilon)$ distinct elements of $I(A)^{(N)}$ (namely the $f(\alpha_0)$, $\alpha_0 \varepsilon F_2$ and $J_1(N, \epsilon)$ corresponding subsets of $I(B)^{(N)}$ (namely the $E(\alpha_0)$) such that (17) holds. Thus if $g < h$, it is possible to transmit at rate G, since, for sufficiently small ϵ ($< h - G$), $J_1(N, \epsilon) > 2^{NG}$ for all sufficiently large N. It now follows from Theorem 4 that $h \leqq H$, and the proof is complete.

REFERENCES

[1] L. Breiman, "The individual ergodic theorem of information theory," *Ann. Math. Stat.*, Vol. 28, No. 3 (1957), pp. 809–811.
[2] A. Feinstein, "A new basic theorem of information theory," *Trans. IRE PGIT* (1954), pp. 2–22.
[3] A. I. Khinchin, *Mathematical Foundations of Information Theory*, Dover Publications, Inc., 1957.
[4] B. McMillan, "The basic theorems of information theory," *Ann. Math. Stat.*, Vol. 24, No. 2 (1953), pp. 196–219.
[5] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27 (1948), pp. 379–423, and 623–656.
[6] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois Journal of Mathematics*, Vol. 1 (1957), pp. 591–606.

ON THE LIMITING POWER FUNCTION OF THE FREQUENCY CHI-SQUARE TEST[1]

By Sujit Kumar Mitra

University of North Carolina[2]

1. Introduction. Several authors have recently investigated the power function of the frequency χ^2-test. Eisenhart [1] and Patnaik [2] have obtained large sample expressions for the power of the simple goodness of fit χ^2-test (i.e. where the class probabilities are completely specified by the null hypothesis). The more complicated case, in which the parameters occurring in the expression for class probabilities require to be estimated, has not received a unified treatment, although the problem has been treated in a number of specific situations by different authors, including, Patnaik [3], Sillito [4], Stevens [5], Pearson and Merrington [6], Poti [7], Chiang [8] and Taylor [9]

Due to difficulties in obtaining the power function of the frequency χ^2-test in the usual manner, Cochran, in an expository article [10] has suggested the derivation of its Pitman limiting power [11], and he illustrated it in the case of the simple goodness of fit test. The concept of asymptotic power suggested by Pitman has also been extensively used in various other areas like nonparametric inference (see e.g. Hoeffding and Rosenblatt [12]) and seems to be a useful tool for comparing alternative consistent tests or alternative designs for experimentation, with regard to their performance in the immediate neighbourhood of the null hypothesis.

The consistency of the frequency χ^2-test has already been established by Neyman [13]. The object of the present paper is to obtain the Pitman limiting power of this test when the unknown parameters occurring in the specification of class probabilities are estimated from the sample by an asymptotically efficient method like the method of maximum likelihood, minimum χ^2 etc. In section 5, we discuss a few applications of the Pitman limiting power for frequency χ^2-tests.

2. Pitman's concept of limiting power [11]. Let H_0 be a certain hypothesis and $\mathfrak{J}$ a test-procedure for testing H_0, which determines the critical region w_n in R_{N_n} (the sample space of N_n dimensions), for $n = 1, 2, \cdots$, ad. inf. Let us assume further that

$$N_{n+1} > N_n \qquad \text{for all } n, \tag{2.1}$$

$$0 < \lim_{n\to\infty} \text{Prob}\,\{w_n \mid H_0\} = \alpha < 1, \tag{2.2}$$

Received July 6, 1956; revised January 7, 1958.

[1] This work was supported in part of the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

[2] Present address: Indian Statistical Institute, Calcutta—35.

and for any alternative H

$$\lim_{n\to\infty} \text{Prob } \{w_n \mid H\} = 1 \tag{2.3}$$

Let $\{H_{0n}\}$ be a family of alternative hypotheses such that

$$\lim_{n\to\infty} \text{Prob } \{w_n \mid H_{0n}\} = \beta(\mathfrak{J}, \{H_{0n}\}) \tag{2.4}$$

exists and $0 < \beta(\mathfrak{J}, \{H_{0n}\}) < 1$.

We call $\beta(\mathfrak{J}, \{H_{0n}\})$ the limiting power of $\mathfrak{J}$ with respect to the family of alternatives $\{H_{0n}\}$.

This concept of limiting power derives its usefulness from the fact that, if $\mathfrak{J}'$ is any other test procedure, which suggests critical regions w_n', instead of w_n, with w_n' satisfying (2.2) and (2.3), and if

$$\beta(\mathfrak{J}, \{H_{0n}\}) \lessgtr \beta(\mathfrak{J}', \{H_{0n}\})$$

then for n sufficiently large

$$\text{Prob } \{w_n \mid H_{0n}\} \lessgtr \text{Prob } \{w_n' \mid H_{0n}\}.$$

3. A theorem in frequency chi-square. Suppose that we have $R = \sum_{i=1}^q r_i$ functions $p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s)$, $(i = 1, 2, \cdots, q; j = 1, 2, \cdots, r_i)$, of $s < R - q$ parameters $\alpha_1, \alpha_2, \cdots, \alpha_s$ such that for all points of a non-degenerate interval A in the s-dimensional space of the α_k's the p_{ij} satisfy the following conditions

(a) $\sum_{j=1}^{r_i} p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s) = 1$ for $i = 1, 2, \cdots, q$,

(b) $p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s) > c^2 > 0$ for all ij,

(c) Every p_{ij} has continuous derivatives $\dfrac{\partial p_{ij}}{\partial \alpha_k}$ and $\dfrac{\partial^2 p_{ij}}{\partial \alpha_k \partial \alpha_t}$,

(d) The matrix $D = \left\{\dfrac{\partial p_{ij}}{\partial \alpha_k}\right\}_{R\times s}$ is of rank s.

(We shall assume that the index pairs (i, j), indicating the rows of the above matrix or of any such matrix we define in future, are arranged in the lexicographic order.)

For $n = 1, 2, \cdots$, ad. inf., let $(N_1^{(n)}, N_2^{(n)}, \cdots, N_q^{(n)})$ be a sequence of row vectors such that for $i = 1, 2, \cdots, q$, and every n, (i) $N_i^{(n)}$ is a natural number, (ii) $N_i^{(n+1)} > N_i^{(n)}$, (iii) if $N_n = \sum_{i=1}^q N_i^{(n)}$, then $N_i^{(n)}/N_n = Q_i$ independent of n.

Let $\alpha_0' = (\alpha_1^0, \alpha_2^0, \cdots, \alpha_s^0)$ be an inner point of A and let

$$c_{ij} (i = 1, 2, \cdots, q; j = 1, 2, \cdots, r_i)$$

be a given set of numbers such that

$$\sum_{j=1}^{r_i} c_{ij} = 0, \qquad \text{for } i = 1, 2, \cdots, q. \tag{3.1}$$

Put

$$p^0_{ij} = p_{ij}(\alpha^0_1, \alpha^0_2, \cdots, \alpha^0_s) \tag{3.2}$$

and

$$p_{ijn} = p^0_{ij} + \frac{c_{ij}}{\sqrt{N_n}} \tag{3.3}$$

Let n_0 be a positive integer such that for $n \geqq n_0$

$$p_{ijn} > 0 \qquad \text{for all } i, j.$$

For $n = n_0, n_0 + 1, \cdots$, ad. inf., let $\{v_{ijn}\}(i = 1, 2, \cdots, q, j = 1, 2, \cdots, r_i)$ be a sequence of R-dimensional random variables such that

$$\text{Prob } \{v_{ijn}\} = \prod_{i=1}^{q} \frac{N_i^{(n)}!}{\prod_{j=1}^{r_i} v_{ijn}!} \prod_{j=1}^{r_i} p_{ijn}^{v_{ijn}}, \tag{3.4}$$

if v_{ijn} are any set of non-negative integers (some of which might be zero) and

$$\sum_{j=1}^{r_i} v_{ijn} = N_i^{(n)}, i = 1, 2, \cdots q,$$

$$= 0, \text{ otherwise.}$$

Consider the system of equations:

$$\sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{v_{ijn} - N_n Q_i p_{ij}}{p_{ij}} \frac{\partial p_{ij}}{\partial \alpha_k} = 0, \qquad k = 1, 2, \cdots, s. \tag{3.5}$$

We shall prove

THEOREM 3.1.

(i) *The system of equations* (3.5) *have exactly one system of solutions*

$$\hat{\alpha}'_n = (\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn})$$

such that $\hat{\alpha}'_n$ *converges in probability to* $\hat{\alpha}'_0$ *as* $n \to \infty$ (or, in symbols, $\hat{\alpha}_n \underset{p}{\to} \alpha_0$ as $n \to \infty$).

(ii) *The value of* χ^2 *obtained by inserting* $\alpha_k = \hat{\alpha}_{kn}$ *in*

$$\chi^2 = \sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{(v_{ijn} - N_n Q_i p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s))^2}{N_n Q_i p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s)} \tag{3.6}$$

is, in the limit as $n \to \infty$, distributed in a non-central χ^2-distribution ([2], [14]), with $R - s - q$ degrees of freedom and non-centrality parameter

$$\lambda = \delta'[I - B(B'B)^{-1}B']\delta,$$

where

$$\delta = \left\{\frac{c_{ij}\sqrt{Q_i}}{\sqrt{p^0_{ij}}}\right\}_{R \times 1},$$

and

$$B = \left\{ \frac{\sqrt{Q_i}}{\sqrt{p_{ij}^0}} \left(\frac{\partial p_{ij}}{\partial \alpha_k} \right)_{\alpha' = \alpha_0'} \right\}_{R \times s}$$

Proof of (i). We observe that for $\eta > d = \max_{i,j} Q_i \mid c_{ij} \mid$,

$$\mid v_{ijn} - N_n Q_i p_{ij}^0 \mid \geqq \eta\sqrt{N_n} \Rightarrow \mid v_{ijn} - N_n Q_i p_{ijn} \mid \geqq (\eta - Q_i \mid c_{ij} \mid)\sqrt{N_n}.$$

Hence, using Chebyshev's inequality, we get

$$\text{Prob}\,\{\mid v_{ijn} - N_n Q_i p_{ij}^0 \mid \geqq \eta\sqrt{N_n}\} \leqq \frac{p_{ijn}(1 - p_{ijn})Q_i}{(\eta - Q_i \mid c_{ij} \mid)^2} < \frac{Q_i p_{ijn}}{(\eta - d)^2}$$

Consequently, the probability that we have $\mid v_{ijn} - N_n Q_i p_{ijn} \mid \geqq \eta\sqrt{N_n}$ for at least one subscript (i, j), is smaller than $(\eta - d)^{-2} \sum_i Q_i \sum_j p_{ijn} = (\eta - d)^{-2}$. Thus with a probability greater than $1 - (\eta - d)^{-2}$, we have

$$\mid v_{ijn} - N_n Q_i p_{ij}^0 \mid < \eta\sqrt{N_n} \qquad \text{for all } (i, j)$$

If we put

$$x_{ijn} = \frac{v_{ijn} - N_n Q_i p_{ij}^0}{\sqrt{N_n Q_i p_{ij}^0}}$$

and $a^2 = \min Q_i$, this will imply that with a probability greater than

$$1 - (\eta - d)^{-2},$$

we have

$$\mid x_{ijn} \mid < \frac{\eta}{ac} \qquad \text{for all } (i, j). \tag{3.7}$$

The proof of Theorem 3.1 (i) can now be completed using (3.7), as well as assumptions (a), (b), (c), and (d) and following Cramér's argument ([15], section 30.3).

Proof of (ii). We put

$$y_{ijn} = \frac{v_{ijn} - N_n Q_i p_{ij}(\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn})}{\sqrt{N_n Q_i p_{ij}(\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn})}}$$

$$X_{(n)} = \{x_{ijn}\}_{R \times 1}$$

$$Y_{(n)} = \{y_{ijn}\}_{R \times 1}$$

$$Z_{(n)} = \{z_{ijn}\}_{R \times 1} = Y_{(n)} - [I - B(B'B)^{-1}B']X_{(n)}$$

The proof of Theorem 3.1 (ii) requires the following results.

Lemma 3.1.

$$z_{ijn} \xrightarrow[p]{} 0 \quad \text{as} \quad n \to \infty$$

Lemma 3.1 can be proved in a manner similar to the proof given in section 30.3 of Cramér's book [15].

LEMMA 3.2. *The limiting distribution of $X'_{(n)}$ is multivariate normal with mean. δ' and covariance matrix*

$$\Lambda_X = I - PP'$$

where

$$P = \{p^0_{ij}\delta_{il}\}_{R\times q}, \quad (i = 1, 2, \cdots, q, \; j = 1, 2, \cdots, n_i), (l = 1, 2, \cdots, q)$$

and δ_{il} is the Kronecker's symbol.

A proof of Lemma 3.2 could be constructed again, on lines similar to that in [15] section 30.1 (see also [16] p 118)

LEMMA 3.3. (Cramér's proposition 22.6 [15]). *Suppose that we have for $v = 1, 2, \cdots$*

$$y_v = Ax_v + z_v ,$$

where x_v, y_v and z_v are n-dimensional random variables, while A is a matrix of order $n \cdot n$ with constant elements. Suppose further that, as $v \to \infty$, the distribution of x_v tends to a certain limiting distribution, while z_v converges in probability to zero. Then y_v has the limiting distribution defined by the linear transformation $y = Ax$, where x has the limiting distribution of the x_v.

LEMMA 3.4. *The limiting distribution of $Y'_{(n)}$ is multivariate normal with mean*

$$\delta'[I - B(B'B)^{-1}B']$$

and covariance matrix

$$\begin{aligned}\Lambda_Y &= [I - B(B'B)^{-1}B'][I - PP'][I - B(B'B)^{-1}B'] \\ &= I - B(B'B)^{-1}B' - PP' \text{ (since } B'P = 0 \text{ as may be verified).}\end{aligned}$$

Lemma 3.4 is a direct consequence of the previous lemmas.

LEMMA 3.5. *There exists an orthogonal matrix L of order $R \cdot R$ such that*

$$L'(I - B(B'B)^{-1}B - PP')L = \begin{array}{c} \\ s+q \\ R-s-q \end{array}\begin{pmatrix} s+q & R-s-q \\ 0 & 0 \\ 0 & I \end{pmatrix}$$

To prove Lemma 3.5, we write

$$M(R \times 2R) = [B(B'B)^{-1}B' \,\vdots\, PP']$$

and observe that

$$B(B'B)^{-1}B' + PP' = MM'$$

Since Rank $[B(B'B)^{-1}B] = s$, Rank $[PP'] = q$ and $B'P = 0$ it follows that Rank $[M] = s + q$. Hence Rank $[B(B'B)^{-1}B + PP'] = s + q$. But

$$B(B'B)^{-1}B + PP'$$

is an idempotent matrix. Hence its only nonzero latent root is 1, which is thus of multiplicity $s + q$. Therefore, since $B(B'B)^{-1}B' + PP'$ is a symmetric matrix, there exists an orthogonal matrix

$$L = [\underset{s+q}{L_1} \vdots \underset{R-s-q}{L_2}]R$$

such that

$$L'(B(B'B)^{-1}B' + PP')L = \begin{matrix} & s+q & R-s-q \\ s+q & I & 0 \\ R-s-q & 0 & 0 \end{matrix}$$

The same matrix L satisfies Lemma 3.5.

If we now make an orthogonal transformation

$$W'_{(n)} = (w_{1,n} w_{2,n}, \cdots, w_{R,n}) = Y'_{(n)}L$$

it will then follow that the limiting distribution of $W'_{(n)}$ is multivariate normal with mean

$$\theta' = \delta'[I - B(B'B)^{-1}B']L$$

and covariance matrix

$$\Lambda_w = \begin{matrix} & s+q & R-s-q \\ s+q & 0 & 0 \\ R-s-q & 0 & I \end{matrix}$$

But

$$B(B'B)^{-1}B + PP' = [L_1 \; L_2] \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \begin{bmatrix} L_1' \\ L_2' \end{bmatrix} = L_1 L_1'$$

Therefore

$$\begin{aligned} I - B(B'B)^{-1}B' &= I - L_1L_1' + PP' \\ &= L_2L_2' + PP', \quad \text{since} \quad LL' = L_1L_1' + L_2L_2' = I \end{aligned}$$

and

$$\begin{aligned} \theta' &= \delta'[I - B(B'B)^{-1}B']L \\ &= \delta'[L_2L_2' + PP'][L_1 \vdots L_2] \\ &= \delta'[PP'L_1 \vdots L_2 + PP'L_2] \\ &= \delta'[0 \vdots L_2], \quad \text{since} \quad \delta'P = 0 \end{aligned}$$

Thus as $n \to \infty$,

$$w_{i,n} \underset{p}{\to} 0, \text{ for } i = 1, 2, \cdots, \overline{s + q},$$

and $w_{s+q+1,n}, w_{s+q+2,n}, \cdots, W_{R,n}$ are asymptotically distributed as independent normal variates with unit variance and means given by

$$\lim_{n \to \infty} E(w_{s+q+1,n}, w_{s+q+2,n}, \cdots, w_{R,n}) = \delta' L_2$$

Hence

$$\sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{(v_{ijn} - N_n Q_i p_{ij} (\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn}))^2}{N_n Q_i p_{ij} (\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn})}$$

$$= Y'_{(n)} Y_{(n)} = W'_{(n)} W_{(n)} = \sum_{i=1}^{R} w_{i,n}^2$$

is, in the limit as $n \to \infty$, distributed as non-central χ^2 with $R - s - q$ degrees of freedom and noncentrality parameter

$$\begin{aligned} \lambda &= \delta' L_2 L_2' \delta \\ &= \delta'(PP' + L_2 L_2')\delta \\ &= \delta'(I - B(B'B)^{-1}B')\delta \end{aligned}$$

This completes the proof of Theorem 3.1 (ii). It will be seen that the proof of Theorem 3.1 given here, follows reasoning similar to that in Cramér ([15], section 30.3). An alternative proof is also possible on the lines of Wald's derivation (Theorem IX [17]) of the large sample distribution of the likelihood ratio criterion, with suitable modifications.

4. The limiting power of the frequency χ^2-test. Neyman [13] considers the following problem:

Consider q sequences of independent trials and let $N_{(i)}$ denote the number of trials in the ith sequence. Each trial of the ith sequence is capable of producing one of the r_i mutually exclusive results, say

$$\rho_{i,1}, \quad \rho_{i,2}, \quad \cdots, \quad \rho_{i,r_i}$$

with unknown probabilities

$$\dot{p}_{i1}, \quad \dot{p}_{i2}, \quad \cdots, \quad \dot{p}_{ir_i}$$

where

$$\sum_{i=1}^{r_i} \dot{p}_{ij} = 1$$

Denote by v_{ij} the number of occurences of $\rho_{i,j}$ in the course of the $N_{(i)}$ trials forming the ith sequence.

On the basis of these observations $\{v_{ij}\}$ it is desired to test the hypothesis

that these unknown probabilities p_{ij}^* satisfy certain known functional relations, e.g.

$$H: p_{ij}^* = p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s)$$

where the p_{ij}'s are certain functions satisfying the conditions described in section 3, and $(\alpha_1, \alpha_2, \cdots, \alpha_s)$ is an unknown parameter point. Let $\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_s$ be a suitably chosen solution of

$$(4.1) \qquad \sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{v_{ij} - N_{(i)} p_{ij}}{p_{ij}} \frac{\partial p_{ij}}{\partial \alpha_k} = 0, \qquad k = 1, 2, \cdots, s.$$

and let $\chi^2_{1-\alpha}(u)$ be the upper α percent point of the χ^2-distribution with u degrees of freedom.

For testing H we compute

$$\chi_H^2 = \sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{(v_{ij} - N_{(i)} p_{ij} (\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_s))^2}{N_{(i)} p_{ij} (\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_s)^2}$$

We reject H if $\chi_H^2 > \chi^2_{1-\alpha}(R - s - q)$, and accept otherwise. Put $N = \sum_{i=1}^{q} N_{(i)}$ and $Q_i = N_{(i)}/N$. Let $\{c_{ij}\}$, $\mathbf{\delta}$ and B be as defined in section 3. Let $F(\chi^2, u, \lambda)$ be the distribution function of the non-central χ^2 with u degrees of freedom and non-centrality parameter λ. Define the hypothesis

$$H_N : p_{ij}^* = p_{ij} (\alpha_1^0, \alpha_2^0, \cdots, \alpha_s^0) + \frac{c_{ij}}{\sqrt{N}} = p_{ijN} \text{ (say)},$$

where as before $(\alpha_1^0, \alpha_2^0, \cdots, \alpha_s^0)$ is an inner point of A.

From Theorem 3.1, we obtain the limiting power of the χ_H^2-test

$$\beta(\chi_H^2, \{H_N\}) = 1 - F(\chi^2_{1-\alpha}(R - s - q), R - s - q, \lambda)$$

where $\lambda = \mathbf{\delta}'(I - B(B'B)^{-1}B)\mathbf{\delta}$.

Let $\mathbf{d}' = (d_1, d_2, \cdots, d_s)$ be any vector of real numbers. When

$$\{c_{ij}\}_{R\times 1} = D\mathbf{d},$$

it is easily seen that

$$p_{ij} (\alpha_1^0, \alpha_2^0, \ldots, \alpha_s^0) + \frac{c_{ij}}{\sqrt{N}} = p_{ij}(\alpha_{1N}, \alpha_{2N}, \ldots, \alpha_{sN}) + o\left(\frac{1}{\sqrt{N}}\right)$$

where $\alpha_{kN} = \alpha_k^0 + d_k/N^{\frac{1}{2}} (k = 1, 2, \cdots, s)$. In this case $\mathbf{\delta}$ is of the form

$$\mathbf{\delta} = B \cdot \mathbf{e}$$

where $\mathbf{e}' = (e_1, e_2, \cdots, e_s)$ is another real vector. We have

$$\lambda = \mathbf{e}'B'(I - B(B'B)^{-1}B')B\mathbf{e}$$
$$= 0$$

and $\beta(\chi_H^2, \{H_N\}) = \alpha$, as we might expect.

5. Applications. (1) *Planning of experiments for comparing two distribution functions.*

To test the hypothesis that two random variables x_1 and x_2 have identical probability distributions, the test procedure commonly adopted consists in making a sequence of N_i independent observations on the random variable x_i $(i = 1, 2)$. At each observation we observe the numerical value assumed by the random variable and according to this classify the results of each sequence into r measurable mutually exclusive and exhaustive groups (same for both the sequences).

Let v_{ij} denote the number of observations of the ith sequence belonging to the jth group $(i = 1, 2, j = 1, 2, \cdots, r)$, so that $\sum_{j=1}^{r} v_{ij} = N_i (i = 1, 2)$. The hypothesis desired to be tested is equivalent to the hypothesis H^* that there are r positive constants $p_1, p_2, \cdots, p_r$ with $\sum_{j=1}^{r} p_j = 1$ such that the probability of a random observation belonging to the jth group is equal to p_j for both the sequences. (We assume that the groups are so chosen that each of them has a positive probability measure at least w.r.t. one of the distributions.)

If this hypothesis H^* is true, the maximum likelihood estimates of p_j will be given by $\hat{p}_j = v_{.j}/N$, where $v_{.j} = v_{1j} + v_{2j}$ and $N = N_1 + N_2$. Hence for testing the hypothesis we compute

$$\chi^2_{H^*} = \sum_{i=1}^{2} \sum_{j=1}^{r} \frac{(v_{ij} - Q_i v_{.j})^2}{Q_i v_{.j}} \tag{5.1}$$

We reject the hypothesis if

$$\chi^2_{H^*} > \chi^2_{1-\alpha}(r - 1),$$

and accept it otherwise.

Let us now assume that it costs C_i dollars to make an observation on x_i $(i = 1, 2)$. Since both N_1 and N_2 are at our disposal it seems now natural to inquire how best we could allocate our total sampling budget of S dollars to the two populations, or, more precisely, could we determine the ratio $N_1 / (N_1 + N_2) = Q_1$ which will maximize the power of the above test with respect to all alternatives violating the hypothesis H^*, and at the same time ensure that the sampling cost does not exceed S dollars. Due to reasons already stated earlier in this paper, we cannot provide an answer to this question with our existing knowledge. However, if we agree to accept the limiting power function as our criterion for choosing 'the best', we might seek if the best possible sampling plan exists in the sense of maximizing the limiting power.

Let c_{ij} $(i = 1, 2, j = 1, 2, \cdots, r)$ be any given set of deviation parameters such that

$$\sum_{j=1}^{r} c_{ij} = 0, \qquad i = 1, 2, \text{ and for at least one } j,$$

$$c_{1j} \neq c_{2j}.$$

Let us denote by H_s^* the hypothesis

$$H_s^* : p_{ij}(S) = p_j^0 + \frac{c_{ij}}{\sqrt{S}}.$$

If we decide to take N_1 and N_2 in the ratio $Q_1 : (1 - Q_1)$ then the total sample size will be given by

$$N = \frac{S}{C_1 Q_1 + C_2 Q_2} \qquad \text{where} \quad Q_2 = 1 - Q_1 .$$

Hence H_s^* may be rewritten as

$$H_s^* : p_{ij}(S) = p_j^0 + \frac{c_{ij}}{\sqrt{C_1 Q_1 + C_2 Q_2}\sqrt{N}}.$$

From Theorem 3.1, we obtain the limiting power of the $\chi^2_{H^*}$-test

$$\beta(\chi^2_{H^*}, \{H_s^*\}) = 1 - F(\chi^2_{1-\alpha}(r-1), (r-1), \chi_{H^*})$$

where

$$\lambda_{H^*} = \delta'(I - B(B'B)^{-1}B')\delta.$$

After some simplification λ_{H^*} reduces to

$$\frac{Q_1 Q_2}{C_1 Q_1 + C_2 Q_2} \sum_{j=1}^{r} (c_{1j} - c_{2j})^2 / p_j^0$$

Since for given x and u, $F(x, u, \lambda)$ is a strictly monotonic decreasing function of λ, the maximum limiting power is attained when λ_{H^*} is maximum, that is when

$$Q_1 = \frac{\sqrt{C_2}}{\sqrt{C_1} + \sqrt{C_2}}$$

Thus to maximize the limiting power the best possible sampling plan, at the specified budget, is given by

$$N_1 = \left[\frac{S}{\sqrt{C_1}\,(\sqrt{C_1} + \sqrt{C_2})}\right]$$

and

$$N_2 = \left[\frac{S}{\sqrt{C_2}\,(\sqrt{C_1} + \sqrt{C_2})}\right]$$

where $[x]$ denotes the largest integer less than x.

(2) *Planning of experiments to detect shifts in response.*

Consider the following problem discussed by McNemar [18] who was interested in ascertaining the effectiveness of an interpolated experience like a movie or a lecture in shifting individual responses to certain stimuli. Let us take the simple

situation in which every individual responds to the stimuli in one of two different ways (say, '0' or '1'). Let π_{ij} denote the proportion of individuals in the population, who give response 'i' before the interpolated experience and response 'j' after it ($i = 0, 1; j = 0, 1$).

Write

$$\pi_{i.} = \pi_{i0} + \pi_{i1} \qquad (i = 0, 1)$$

$$\pi_{.j} = \pi_{0j} + \pi_{1j} \qquad (j = 0, 1)$$

We shall say there is no shift in response if

$$H_0 : \pi_{1.} = \pi_{.1}$$

is true.

To test this hypothesis one can conceive of at least two alternative ways of experimentation:

(a) two samples, each of size n are selected independently, one from the pre-experience group and the other from the post-experience group. The test for the equality of proportions then, is easily seen to be a particular case of the test given earlier in this section under Application (1) Let us denote the chisquare obtained for this test by χ_a^2.

(b) the same set of individuals, n, in number selected from the pre-experience group is again examined after the experience, and the results classified in a 2 × 2 table as follows:

	Post experience response		
	0	1	total
Pre-experience response			
0	n_{00}	n_{01}	$n_{0.}$
1	n_{10}	n_{11}	$n_{1.}$
total	$n_{.0}$	$n_{.1}$	n

Under procedure (b), to test H_0, we compute

$$\chi_b^2 = \frac{(n_{10} - n_{01})^2}{n_{01} + n_{10}}$$

and reject H_0, only if, $\chi_b^2 > \chi_{1-\alpha}^2\,(1)$. Let us denote by H_{0n} the hypothesis

$$H_{0n} : \pi_{ij} = \pi_{ij}^0 + \frac{c_{ij}}{\sqrt{n}},$$

where $\Sigma\pi_{ij}^0 = 1$, $\pi_{01}^0 = \pi_{10}^0 = \pi^0$ (say), $\Sigma c_{ij} = 0$, and $c_{01} \neq c_{10}$. From Theorem 3.1, we obtain after certain algebraic simplification:

$$\beta(\chi_a^2, \{H_{0n}\}) = 1 - F(\chi_{1-\alpha}^2(1), 1, \lambda_a)$$

and

$$\beta(\chi_b^2, \{H_{0n}\}) = 1 - F(\chi_{1-\alpha}^2(1), 1, \lambda_b),$$

where

$$\lambda_a = \frac{(c_{10} - c_{01})^2}{2\{(\pi_{11}^0 + \pi^0)(\pi_{00}^0 + \pi^0)\}}$$

and

$$\lambda_b = \frac{(c_{10} - c_{01})^2}{2\pi^0}.$$

The denominator in λ_a can be rewritten as $2\{\pi^0 - \pi^{0^2} + \pi_{00}^0\pi_{11}^0\}$. Hence, $\lambda_b >, <$ or $= \lambda_a$, according as $(\pi_{00}^0\pi_{11}^0 - \pi_{01}^0\pi_{10}^0) >, <$ or $= 0$ respectively. This shows that at least from the point of view of maximising limiting power, procedure (b) would be superior to procedure (a) when the association between the two response types, as measured by $\pi_{00}^0\pi_{11}^0 - \pi_{01}^0\pi_{10}^0$, is positive; inferior to (a) when it is negative; and equivalent to (a) when it is zero.

ACKNOWLEDGEMENT. I would like to thank Professor S. N. Roy for his continued interest in this work and his helpful suggestions, and the referees for their valuable comments.

REFERENCES

[1] C. EISENHART, "The power function of the χ^2 test," (Abstract) *Bull. Amer. Math. Soc.*, Ser. 2, Vol. 44 (1938), p. 32.
[2] P. B. PATNAIK, "The non-central χ^2 and F-distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202–232.
[3] P. B. PATNAIK, "The power function of the test for the difference between two proportions in a 2 × 2 table," *Biometrika*, Vol. 35 (1948), pp. 157–175.
[4] G. P. SILLITTO, "Note on approximations to the power function of the '2 × 2 comparative trial'," *Biometrika*, Vol. 36 (1949), pp. 347–352.
[5] W. L. STEVENS, "Mean and variance of an entry in a contingency table," *Biometrika*, Vol. 38 (1951), pp. 468–470.
[6] E. S. PEARSON AND M. MERRINGTON, "2 × 2 Tables; the power function of the test on a randomized experiment," *Biometrika*, Vol. 35 (1948), pp. 331–345.
[7] S. J. POTI, "Power function of Chi-Square test with special reference to analysis of blood group data," *Sankhya*, Vol. 10 (1950), pp. 397–406.
[8] C. L. CHIANG, "On the Design of Mass Medical Surveys," *Human Biology*, Vol. 23 (1951), pp. 242–271.
[9] W. F. TAYLOR, "Mathematical Statistical Studies on the Tuberculin and Histoplasmin skin tests, I, overlapping of Negative and Positive Reactions," *Human Biology*, Vol. 23 (1951), pp. 1–23.
[10] W. G. COCHRAN, "The χ^2 test of goodness of fit," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 314–345.
[11] E. J. G. PITMAN, "Unpublished lecture notes on non-parametric statistical inference," 1948.
[12] W. HOEFFDING AND J. R. ROSENBLATT, "The Efficiency of Tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 52–63.
[13] J. NEYMAN, "Contributions to the theory of χ^2 test," *Proceedings of the Berkeley Sym-*

posium on Mathematical Statistics and Probability, University of California Press, 1949.

[14] E FIX, "Tables of Noncentral χ^2," *University of California Publications in Statistics*, Vol. 1, No. 2, pp. 15–19, University of California Press, 1949.

[15] H CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946

[16] S K. MITRA, *Contributions to the Statistical Analysis of Categorical Data*, Institute of Statistics Mimeograph Series, University of North Carolina, No. 142 (1955), pp. 1–146.

[17] A WALD, "Tests of Statistical Hypotheses Concerning Several Parameters when the number of observations is large," *Trans Amer. Math Soc*, Vol. 54, (1943), pp. 426–482.

[18] Q MCNEMAR, *Psychological Statistics*, 2nd edition, Chapman and Hall, London, 1955.

SOME EXACT RESULTS FOR THE FINITE DAM

By N. U. Prabhu

University of Western Australia, and Karnatak University, India

1. Summary. In the discrete finite dam model due to Moran, the storage process $\{Z_t\}$ is known to be a Markov chain. Stationary distributions of Z_t are obtained for the cases where the release is a unit amount of water per unit time, and the input is of (i) geometric, (ii) negative binomial and (iii) Poisson type.

The paper concludes with a discussion of the problem of emptiness in the finite dam and considers the probability that, starting with an arbitrary storage, the dam becomes empty before it overflows.

2. Introduction. This paper is concerned with a storage system whose probability model is due to Moran [9]. The storage Z_t of a dam of finite capacity K is defined for discrete time t ($t = 0, 1, 2, \cdots$) as the dam content just after an instantaneous release at t, and just before an input X_t flows into it over the time-interval $(t, t + 1)$. The model is subject to the conditions that

(i) the inputs X_t during the intervals $(t, t + 1)$ are independently and identically distributed;

(ii) there is an overflow $\text{Max}(Z_t + X_t - K, 0)$ during the interval $(t, t + 1)$, a quantity $\text{Min}(K, Z_t + X_t)$ being left in the dam just before the release occurs; and

(iii) the amount of water released at time $t + 1$ is $\text{Min}(M, Z_t + X_t)$ where M is a constant $(<K)$.

A fuller description of the model and further references on the subject are given by Gani [3]. It is seen the stochastic processes $\{Z_t\}$ and $\{Z_t + X_t\}$ are both Markov chains, and the problem of obtaining their stationary distributions, given the probability distribution of the input, is of some interest. Moran ([9], [10]) and Gani and Moran [4] have obtained a few approximate solutions to this problem by numerical methods, and some important observations on the solution in the general case have been made by Moran [11], but the only exact solution known so far is the one due to Moran [10] for the case of the geometric input. The problem is considerably simplified when $K = \infty$, i.e. when the dam is of infinite capacity; it is then seen (Gani and Prabhu, [5]) that the transition-matrix of the Markov chain $\{Z_t + X_t\}$ also occurs in the theory of queues in connection with the length of a queue at epochs just before service. For this case Bailey [1] has obtained, by the method of probability generating functions (p.g.f.), the stationary distributions arising from a given distribution of X_t. A dam of finite capacity K can be considered as the analogue of a queueing system in which there is accommodation for only K customers to wait, those in excess of K being compelled to leave the queue altogether (as may happen, for

Received November 8, 1957; revised March 27, 1958.

instance, in an airport of limited capacity); we proceed to obtain for such a dam the stationary distribution of the storage Z_t.

3. Stationary distribution of the storage. We shall be concerned with the case where M, the amount of water released at time t, is unity. Let $\{g_j\}$ be the probability distribution of X_t, so that

$$\Pr\{X_t = j\} = g_j, \qquad (j = 0, 1, 2, \cdots). \tag{1}$$

We assume that $g_j > 0$ for all j. Also, let

$$G(z) = \sum_{j=0}^{\infty} g_j z^j, \qquad |z| < 1 \tag{2}$$

be the p.g.f. of $\{g_j\}$, and

$$\rho = G'(1) = \sum_{j=0}^{\infty} j g_j \tag{3}$$

the mean input. The transition-matrix of the Markov chain $\{Z_t\}$ is $P \equiv \{P_{ij}\}$, where

$$P = \begin{array}{c|ccccc} \nearrow & 0 & 1 & & K-2 & K-1 \\ \hline 0 & g_0 + g_1 & g_2 & \cdot & g_{K-1} & h_K \\ 1 & g_0 & g_1 & \cdot & g_{K-2} & h_{K-1} \\ 2 & 0 & g_0 & & g_{K-3} & h_{K-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & & & & \cdot \\ K-1 & 0 & 0 & \cdot & g_0 & h_1 \end{array} \tag{4}$$

where $h_i = \sum_{j=i}^{\infty} g_j$, $(i = 1, 2, \cdots, K)$ Clearly, the chain is irreducible and contains a finite number K of states, so that the stationary probability distribution $\{u_i\}$, $(i = 0, 1, \cdots, K-1)$ exists, where the u_i are the unique solutions of the equations

$$u_j = \sum_{i=0}^{K-1} u_i p_{ij}, \qquad (j = 0, 1, \cdots, K-1) \tag{5}$$

together with $\sum_0^{K-1} u_i = 1$. We first prove the following theorem due to Moran [11].

THEOREM.
(i) *If* $\{u_i^{(K)}\}$, $(i = 0, 1, \cdots, K-1)$ *is the stationary probability distribution of storage in a dam of capacity* K, *then the ratios*

$$v_i = \frac{u_i^{(K)}}{u_0^{(K)}}, \qquad (i = 1, 2, \cdots, K-1) \tag{6}$$

are independent of K, *and*

(ii) *the v_i's can be found as the coefficients of z^i in $V(z)$, where*

$$v(z) = \frac{g_0(1-z)}{G(z)-z} \tag{7}$$

The first part of the theorem is easily proved; in fact, writing out the equations (5) in full we obtain

$$\begin{aligned} u_0 &= (g_0+g_1)u_0 + g_0u_1 \\ u_1 &= g_2u_0 \quad + g_1u_1 + g_0u_2 \\ \cdot \quad & \quad \cdot\cdot \qquad \cdot\cdot \qquad \cdot\cdot \\ u_{K-2} &= g_{K-1}u_0 + g_{K-2}u_1 + \cdots + g_0u_{K-1} \\ u_{K-1} &= h_Ku_0 \quad + h_{K-1}u_1 + \cdots + h_1u_{K-1} \end{aligned}$$

Solving these equations successively for the ratios $v_i = u_i / u_0$ we obtain

$$\begin{aligned} v_1 &= \frac{1-g_0-g_1}{g_0} \\ v_2 &= \frac{1-g_1}{g_0}v_1 + \frac{g_2}{g_0} \end{aligned} \tag{8}$$

and in general, the v_i's ($i = 1, 2, \cdots, K-1$) are seen to be independent of K. Now consider the function $V(z)$ defined by (7). We shall first prove that $V(z)$ can be expanded as a power series which is convergent for suitable values of $|z|$. Let us first consider the case $\rho \leqq 1$. Writing

$$G(z) - z = (1-z)\left\{1 - \frac{1-G(z)}{1-z}\right\}$$

and following Kendall ([6], p. 159) we obtain

$$\frac{1-G(z)}{1-z} = \sum_{n=0}^{\infty} z^n \sum_{n+1}^{\infty} g_i$$

so that, for $|z| < 1$,

$$\left|\frac{1-G(z)}{1-z}\right| < \sum_{n=0}^{\infty}\sum_{n+1}^{\infty} g_i = \sum_{i=1}^{\infty} ig_i = \rho \leqq 1$$

Hence $|G(z) - z| \neq 0$, and we have the power series expansion

$$V(z) = g_0\left\{1 - \frac{1-G(z)}{1-z}\right\}^{-1} = v_0 + v_1 z + v_2 z^2 + \cdots$$

convergent for $|z| < 1$.

Next, let $\rho > 1$. In this case there exists a positive λ such that the power series expansion

$$\frac{1}{G(z)-z} = c_0 + c_1 z + c_2 z^2 + \cdots$$

is valid for $|z| < \lambda$ (Knopp, [8], p. 182). Hence it follows that $V(z)$ also possesses a power series expansion convergent for $|z| < \lambda$.

Thus whether or not $\rho \leqq 1$, $V(z)$ has a power series expansion

$$V(z) = \frac{g_0(1 - z)}{G(z) - z} = v_0 + v_1 z + v_2 z^2 + \cdots \tag{9}$$

The coefficients v_i are determined from the relation

$$g_0(1 - z) = \{G(z) - z\} \sum_{i=0}^{\infty} v_i z^i$$

and hence it is seen that $v_0 = 1$, and $v_1, v_2, \cdots, v_{K-1}$ satisfy the relations (8). Thus they are, in fact, the quantities defined in (6).

If $\rho < 1$, the stationary probability distribution exists in the case of the infinite chain ($K = \infty$), and its g.f. is proportional to $V(z)$. However, the above results hold, as we have shown, even when $\rho \geqq 1$. It is now obvious that the general method of obtaining the stationary probability distribution $\{u_i\}$ for the discrete dam of finite capacity K consists of (i) finding $V(z)$, (ii) expanding $V(z)$ to obtain the v_i's, and (iii) normalising $v_0, v_1, \cdots, v_{K-1}$ to obtain a probability distribution. We proceed to do this in some particular cases.

3.1. Geometric input. Consider, for instance, an input distribution of the geometric type,

$$g_j = \Pr\{X_t = j\} = ab^j, \qquad (j = 0, 1, \cdots,) \tag{10}$$

where $0 < a < 1$ and $b = 1 - a$. The p.g.f. of X_t is then

$$G(z) = \frac{a}{1 - bz} \tag{11}$$

and the function $V(z)$ is given by

$$\begin{aligned} V(z) &= \frac{a(1 - z)}{a(1 - bz)^{-1} - z} = \frac{1 - bz}{1 - \rho z} \\ &= (1 - bz) \sum_{i=0}^{\infty} \rho^i z^i \left(|z| < \min\left(\frac{1}{\rho}, 1\right)\right), \end{aligned} \tag{12}$$

where $\rho = b/a$ is the mean input. Hence we obtain

$$v_0 = 1, \qquad v_i = \rho^i - b\rho^{i-1} = b\rho^i, \qquad (i = 1, 2, \cdots, K - 1)$$

and

$$\sum_0^{K-1} v_i = 1 + b \sum_{i=1}^{K-1} \rho^i = a \frac{1 - \rho^{K+1}}{1 - \rho}.$$

The stationary distribution in this case is therefore given by $\{u_i\}$, where

$$u_0 = \frac{(1 - \rho)}{a(1 - \rho^{K+1})}, \qquad u_i = \frac{\rho^{i+1}(1 - \rho)}{1 - \rho^{K+1}}, \qquad (i = 1, 2, \cdots, K - 1), \tag{13}$$

Thus the storage of a dam of finite capacity K into which flows an input of the geometric type has a stationary distribution of the geometric type, which is truncated at $Z = K - 1$ and has a modified initial term. This result is implied in Moran's solution (referred to in Section 2) for the general case $M > 1$, although it is not explicitly mentioned by him; for $M = 1$ his solution is given by the formulae $u_0 = \pi_0 + \pi_1$, $u_i = \pi_{i+1}$ $(i = 1, 2, \cdots, K - 1)$, where

$$\pi_{K-r}/\pi_K = {}^rS_1\, a - {}^{r-1}S_2\, a^2 + {}^{r-2}S_3\, a^3 \cdots$$

(14) $$ {}^nS_p = \binom{n-1}{p-1} b^{-n} - \binom{n-2}{p-1} b^{1-n}. \qquad (r = 1, 2, \cdots, K)$$

From this we obtain (13) after some simple reduction.

3.2. Negative binomial input. Consider next the more general case of the negative binomial input,

(15) $$g_j = \Pr\{X_t = j\} = n_j \binom{n+j-1}{j} a^n b^j, \qquad (j = 0, 1, \cdots)$$

where $0 < a < 1$, $b = 1 - a$, and n is a positive integer; the p.g.f. of X_t is then

$$G(z) = \frac{a^n}{(1 - bz)^n}$$

and the mean input is $\rho = nb/a$. We have then

(16) $$V(z) = \frac{a^n(1 - z)}{a^n(1 - bz)^{-n} - z} = \frac{a^n(1 - z)(1 - bz)^n}{a^n - z(1 - bz)^n}$$

Obviously $z = 1$ is a zero of the denominator of the expression on the right hand side of (16); in addition to this it has n other zeros $z_1, z_2, \cdots, z_n$. We consider here the case where $z_1, z_2, \cdots, z_n$ are all distinct and different from unity; however, the general case can be treated along similar lines. When $(1, z_1, z_2, \cdots, z_n)$ are all different we can break up $V(z)$ into partial fractions of the form

(17) $$V(z) = d_0 + \sum_{p=1}^{n} \frac{d_p}{1 - z/z_p}$$

where obviously $d_0 = a^n$ and the d_p's are given by

$$d_p = \lim_{z \to z_p} \left(1 - \frac{z}{z_p}\right) V(z)$$

(18) $$= \lim_{z \to z_p} \frac{a^n(1 - z)(1 - bz)^n \left(1 - \dfrac{z}{z_p}\right)}{a^n - z(1 - bz)^n} = \frac{a^n(1 - 1/z^p)}{\rho a z_p/(1 - bz_p) - 1}$$

$$(p = 1, 2, \cdots, n).$$

Now let λ be the least among the quantities $1, |z_1|, |z_2|, \cdots, |z_n|$; then for $|z| < \lambda$ we can express each term under the summation sign in (17) as a power series. Thus

$$V(z) = d_0 + \sum_{p=1}^{n} d_p \sum_{i=0}^{\infty} \left(\frac{z}{z_p}\right)^i = d_0 + \sum_{i=0}^{\infty} z^i \sum_{p=1}^{n} d_p \left(\frac{1}{z_p}\right)^i,$$

whence we obtain

$$\begin{aligned} v_0 &= d_0 + \sum_{p=1}^{n} d_p = \lim_{z \to 0} V(z) = 1 \\ v_i &= \sum_{p=1}^{n} \frac{d_p}{(z_p)^i}, \qquad (i = 1, 2, \cdots, K-1), \end{aligned} \tag{19}$$

so that

$$\sum_{i=0}^{K-1} v_i = d_0 + \sum_{p=1}^{n} d_p \sum_{i=0}^{K-1} \left(\frac{1}{z_p}\right)^i = d_0 + \sum_{p=1}^{n} d_p \frac{1 - (1/z_p)^K}{1 - (1/z_p)}.$$

It follows that the stationary probabilities u_i are given by

$$\begin{aligned} u_0 &= \left\{ d_0 + \sum_{p=1}^{n} d_p \frac{1 - (1/z_p)^K}{1 - (1/z_p)} \right\}^{-1} \\ u_i &= u_0 \sum_{p=1}^{n} d_p \left(\frac{1}{z_p}\right)^i, \qquad (i = 1, 2, \cdots K-1). \end{aligned} \tag{20}$$

From (20) we see that the stationary distribution of the dam storage is the weighted sum of n geometric distributions, each of which is truncated at $Z = K - 1$, and has a modified initial term.

3.3. Poisson input. Finally we consider the case where the input has the Poisson distribution with mean ρ,

$$g_j = \Pr\{X_t = j\} = e^{-\rho} \frac{\rho^j}{j!}, \qquad (j = 0, 1, \cdots). \tag{21}$$

The rigorous procedure here consists of writing down $V(z)$ and obtaining the coefficients v_i by complex variable methods. We shall, however, argue heuristically and consider (21) as the limiting case of the negative binomial (15) as $n \to \infty$, $a \to 1$ and $\rho = nb / a$ is held fixed. In fact, putting $a = 1 / (1 + \rho/n)$, $b = \rho/(n + \rho)$, it is seen that the p g.f. of (15) reduces to

$$(1 + \rho/n)^{-n} \left\{ 1 - \frac{1}{n} \rho \left(1 + \frac{1}{n} \rho\right)^{-1} z \right\}^{-n} \to e^{-\rho + \rho z}$$

which is the p.g.f. of (21). Also, $d_0 \to e^{-\rho}$, and

$$d_p = \frac{a_n(1 - 1/z_p)}{\rho a z_p/(1 - b z_p) - 1} \to \frac{e^{-\rho}(1 - 1/z_p)}{\rho z_p - 1}$$

where $z_1, z_2, \cdots$ are the roots (other than unity) of the equation

$$e^{-\rho+\rho z} = z \tag{22}$$

(which are infinite in number). Hence the stationary probabilities of the dam storage are given by

$$\begin{aligned} u_0 &= \left\{e^{-\rho} + \sum_{p=1}^{\infty} \frac{e^{-\rho}(1-1/z_p)}{\rho z_p - 1} \frac{1-(1/z_p)^K}{1-(1/z_p)}\right\}^{-1} \\ u_i &= u_0 \sum_{p=1}^{\infty} \frac{e^{-\rho}(1-1/z_p)}{\rho z_p - 1}\left(\frac{1}{z_p}\right)^i, \qquad (i = 1, 2, \cdots, K-1). \end{aligned} \tag{23}$$

4. The problem of emptiness in the finite dam. The analogy between the dam process and the random walk has already been pointed out by several authors (see the discussion in [3]). In fact, putting $U_t = X_t - 1$, we see that the storage Z_t in a dam of capacity K satisfies the relations

$$Z_{t+1} = \begin{cases} Z_t + U_t & \text{if } 0 < Z_t + U_t < K-1 \\ 0 & \text{if } \quad Z_t + U_t \leqq 0 \\ K-1 & \text{if } \quad Z_t + U_t \geqq K-1 \end{cases} \tag{24}$$

which, however, define a random walk with impenetrable barriers at $Z = 0$ and $Z = K - 1$. If $K = \infty$, there is only the first barrier and the problem of 'duration of the game' (i.e. the distribution of time required for the dam to become empty for the first time) has been discussed by Kendall [7] for the case where the input is of the Gamma type and the release is continuous. For finite K this problem is much more difficult; however, for this case we propose to discuss the probability of absorption at $Z = 0$ (i.e. the conditional probability V_i that, starting with a storage $Z_0 = i$, the dam becomes empty ($Z_t = 0$) before it overflows). This is a familiar problem in random walk theory, and has been discussed, for instance, by Feller ([2], pp. 300–303); it is seen that the probabilities V_i $(i = 1, 2, \cdots, K-2)$ satisfy the relations

$$\begin{aligned} V_1 &= \sum_{j=1}^{K-2} g_j V_j + g_0 \\ V_i &= \sum_{j=i-1}^{K-2} g_{j-i+1} V_j, \qquad (i = 2, 3, \cdots, K-2). \end{aligned}$$

These equations simplify to some extent if we note that the states 0 and $K - 1$ are absorbing, so that $V_0 = 1$, $V_{K-1} = 0$; for we can then write

$$V_i = \sum_{j=i-1}^{K-1} g_{j-i+1} V_j, \qquad (i = 1, 2, \cdots, K-2). \tag{25}$$

Clearly, the coefficients on the right hand side of these equations correspond to the rows of the transition-matrix (4). It will now be found easiest to start at the

bottom right hand corner and work up to the left: thus

$$g_0 V_{K-3} + g_1 V_{K-2} + h_2 \cdot 0 = V_{K-2}$$

so that

$$V_{K-3} = \frac{1 - g_1}{g_0} V_{K-2},$$

and similarly

$$V_{K-4} = \frac{(1 - g_1) V_{K-3} - g_2 V_{K-2}}{g_0},$$

etc. This shows that the ratios of the quantities

$$w_i = V_{K-1-i} \tag{26}$$

are again independent of K ($w_0 = 0$, $w_{K-1} = 1$); rewriting the equations (25) in terms of these quantities we obtain

$$w_i = \sum_{j=0}^{i} g_j w_{i-j+1}, \qquad (i = 1, 2, \cdots, K - 2). \tag{27}$$

Consider the system of equations (27) for $i = 1, 2, \cdots$ ad. inf., and put

$$W(z) = \sum_{i=1}^{\infty} \frac{w_i}{w_1} z^{i-1} \tag{28}$$

we have

$$\begin{aligned} zW(z) &= \sum_{i=1}^{\infty} \frac{z^i}{w_1} \sum_{j=0}^{i} g_j w_{i-j+1} \\ &= \sum_{j=1}^{\infty} g_j \sum_{i=j}^{\infty} \frac{w_{i-j+1}}{w_1} z^i + g_0 \sum_{i=1}^{\infty} \frac{w_{i+1}}{w_1} z^i \\ &= \sum_{j=1}^{\infty} g_j \sum_{i=1}^{\infty} \frac{w_i}{w_1} z^{i+j-1} + g_0 \sum_{i=2}^{\infty} \frac{w_i}{w_1} z^{i-1} \\ &= G(z)W(z) - g_0, \end{aligned}$$

whence we obtain the relation

$$W(z) = \frac{g_0}{G(z) - z}. \tag{29}$$

Folowing the same lines of argument as for $V(z)$, we can prove that $W(z)$ can be expanded as a power series convergent for suitable values of $|z|$. Let $W(z) = \sum_{i=0}^{\infty} \omega_{i+1} z^i$; then since $\omega_{K-1} = w_{K-1} / w_1 = 1/w_1$, we must have

$$w_i = \frac{\omega_i}{\omega_{K-1}}, \qquad (i = 1, 2, \cdots, K - 2) \tag{30}$$

which are, therefore, the required solutions to the equations (27). The absorption probabilities V_i can then be obtained from (26).

Let us now consider the particular case where the input is geometric with probabilities $g_j = ab^j$, $(j = 0, 1, 2, \cdots)$, and $G(z) = a(1 - bz)^{-1}$; then (29) gives

$$W(z) = \frac{a}{a(1-bz)^{-1} - z} = \frac{(1-bz)}{(1-z)(1-\rho z)}$$

(31)
$$= \begin{cases} \dfrac{a}{1-\rho}\left\{\dfrac{1}{1-z} - \dfrac{\rho^2}{1-\rho z}\right\} & \text{if } \rho \neq 1 \\ \dfrac{1-bz}{(1-z)^2} & \text{if } \rho = 1. \end{cases}$$

Hence it follows that

(32)
$$\omega_i = \begin{cases} \dfrac{a(1-\rho^{i+1})}{1-\rho} & \text{if } \rho \neq 1 \\ a(i+1) & \text{if } \rho = 1 \end{cases} \qquad (i \neq 1)$$

and

(33)
$$w_i = \begin{cases} \dfrac{1-\rho^{i+1}}{1-\rho^K} & \text{if } \rho \neq 1 \\ \dfrac{i+1}{K} & \text{if } \rho = 1 \end{cases} \qquad (i = 1, 2, \cdots, K-2).$$

Thus the absorption probabilities V_i in the case of the geometric input are given by

(34)
$$V_i = \begin{cases} \dfrac{1-\rho^{K-i}}{1-\rho^K} & \text{if } \rho \neq 1 \\ 1 - \dfrac{i}{K} & \text{if } \rho = 1 \end{cases} \qquad (i = 1, 2, \cdots, K-2).$$

A similar procedure could be used, when the input is of a more general type, to obtain the exact expressions for the probabilities V_i. However, in many cases, it may suffice to know the bounds within which V_i lie, and these bounds are given by Feller ([2], inequalities 8.11 and 8.12 on p. 303). In fact, noting that $E(U_t) = E(X_t - 1) = \rho - 1$, where ρ is the mean input, we have that

(35)
$$\begin{aligned} \frac{z_0^{K-1} - z_0^i}{z_0^{K-1} - 1} &\leqq V_i \leqq 1 && \text{if } \rho < 1 \\ \frac{z_0^i - z_0^{K-1}}{1 - z_0^{K-1}} &\leqq V_i \leqq z_0^i && \text{if } \rho > 1 \\ 1 - \frac{i}{K-1} &\leqq V_i \leqq 1 && \text{if } \rho = 1 \end{aligned}$$

where z_0 is the unique positive root (other than unity) of the equation $\sum_j z^j \Pr\{U_t = j\} = 1$, i.e. $\sum_{j=0}^{\infty} g_j z^j = z$, and $z_0 \gtrless 1$ according as $\rho \lessgtr 1$.

5. Concluding remarks and acknowledgements. When the input X_t has a continuous probability distribution, it is seen that the stationary distribution function of $Z_t + X_t$ satisfies an integral equation, which has been solved by the author in a recent paper (Prabhu, [12]) for the special case when the input distribution is of the Gamma type. A more realistic problem on which some work is in progress at the moment is the one dealing with the finite dam process in continuous time; however, our solutions for discrete time may be taken as useful approximations to this continuous case.

I am indebted to Dr. J. Gani and the referee for many helpful suggestions.

REFERENCES

[1] N. T. J. Bailey, "On queueing processes with bulk service," *J Roy Stat. Soc*, B, Vol. 16 (1954), pp. 80–87.

[2] W. Feller, "An Introduction to Probability Theory and its Applications," John Wiley and Sons, 1950

[3] J Gani, "Problems in the probability theory of storage systems," *J Roy Stat Soc.*, B, Vol. 14 (1957), pp. 181–207.

[4] J. Gani and P. A. P. Moran, "A solution of dam equations by Monte Carlo methods," *Aust. J. Appl. Sci.*, Vol. 6 (1955), pp 267–273.

[5] J. Gani and N. U. Prabhu, "Stationary distributions of the negative exponential type for the infinite dam," *J. Roy. Stat. Soc*, B Vol. 19 (1957), pp. 342–351.

[6] D. G. Kendall, "Some problems in the theory of queues," *J Roy. Stat Soc*, B Vol. 13 (1951), pp. 151–173.

[7] D. G. Kendall, "Some problems in the theory of dams," *J Roy Stat Soc*, B Vol. 19 (1957), pp. 207–212

[8] K. Knopp, "Theory and applications of infinite series," Blackie and Son, London and Glasgow, 1928

[9] P. A. P. Moran, "A probability theory of dams and storage systems," *Aust J Appl. Sci.*, Vol. 5 (1954), pp. 116–124.

[10] P. A. P. Moran, "A probability theory of dams and storage systems: modifications of the release rules," *Aust. J. Appl Sci*, Vol. 6 (1955), pp. 117–130.

[11] P. A. P. Moran, "A probability theory of dams with a continuous release," *Quart. J. of Math.*, (Oxford, 2), Vol. 7, pp. 130–137

[12] N. U. Prabhu, "On the integral equation for the finite dam," *Quart. J. of Math.*, (Oxford, 2), (to appear).

MINIMAX ESTIMATION FOR LINEAR REGRESSIONS[1]

By R. Radner

University of California, Berkeley

1. Introduction and Summary. When estimating the coefficients in a linear regression it is usually assumed that the covariances of the observations on the dependent variable are known up to multiplication by some common positive number, say c, which is unknown. If this number c is known to be less than some number k, and if the set of possible distributions of the dependent variable includes "enough" normal distributions (in a sense to be specified later) then the minimum variance linear unbiased estimators of the regression coefficients (see [1]) are minimax among the set of all estimators; furthermore these minimax estimators are independent of the value of k. (The risk for any estimator is here taken to be the expected square of the error.) This fact is closely related to a theorem of Hodges and Lehmann ([3], Theorem 6.5), stating that if the observations on the dependent variable are assumed to be independent, with variances not greater than k, then the minimum variance linear estimators corresponding to the assumption of equal variances are minimax.

For example, if a number of observations are assumed to be independent, with common (unknown) mean, and common (unknown) variance that is less than k; and if, for every possible value of the mean, the set of possible distributions of the observations includes the normal distribution with that mean and with variance equal to k; then the sample mean is the minimax estimator of the mean of the distribution.

The assumption of independence with common unknown variance is, of course, essentially no less general than the assumption that the covariances are known up to multiplication by some common positive number, since the latter situation can be reduced to the former by a suitable rotation of the coordinate axes (provided that the original matrix of covariances is non-singular).

This note consideres the problem of minimax estimation, in the general "linear regression" framework, when less is known about the covariances of the observations on the "dependent variable" than in the traditional situation just described. For example, one might not be sure that these observations are independent, nor feel justified in assuming any other specific covariance structure. It is immediately clear that, from a minimax point of view, one cannot get along without any prior information at all about the covariances, for in that case the risk of every estimator is unbounded. In practice, however, one is typically willing to grant that the covarainces are bounded somehow, but one may not

Received June 4, 1957; revised May 5, 1958.

[1] Research undertaken by the Cowles Commission for Research in Economics under Contract Nonr-358(01), NR 047-006 with the Office of Naval Research.

have a very precise idea of the nature of the bound. One is therefore led to look for different ways of bounding the covariances, in the hope that the minimax estimators are not too sensitive to the bound.

Unfortunately, in the directions explored here, the minimax *estimator is* sensitive to the "form" of the bound, although once the form has been chosen the minimax estimator does not depend on the "magnitude" of the bound. This result thus provides an instance in which the minimax principle is not too effective against the difficulties due to vagueness of the statistical assumptions of a problem, although this is a type of situation in which it has often been successful (see Savage in [4], pp. 168–9).

In this note, two ways of bounding the covariances are considered. The first is equivalent to choosing a coordinate system for the "dependent variables," and placing a bound on the characteristic roots of the matrix of covariances of the coordinates, in terms of one of a certain class of metrics (e.g., placing a bound on the trace on the covariance matrix, or on its largest characteristic root). The second way consists of choosing a coordinate system, and then placing a bound on the variance of each coordinate.

In the first situation, the minimum variance linear unbiased estimator corresponding to the case of uncorrelated coordinates, with equal variances, turns out to be minimax; this minimax estimator is, in general, different for different choices of coordinate system, but does not depend on the "magnitude" of the bound. Also, the minimax loss typically decreases at the rate of the reciprocal of the sample size.

In the second situation, the minimax procedures derived here involve ignoring most of the observations, and applying a linear unbiased estimator to the rest. Again, the minimax procedure depends upon the choice of coordinate system; furthermore, in this case the minimax loss typically either does not approach zero with increasing sample size, or does so much more slowly than the reciprocal of the sample size.

Thus the minimax estimator appears to be less unsatisfactory in the first situation than in the second, but in both cases it depends upon the choice of coordinate system, which is a disadvantage if there is no "natural" coordinate system intrinsic to the regression problem being considered.

Section 2 below presents the formulation of the problem, and a basic lemma. Sections 3 and 4 explore the two ways of bounding the covariances just mentioned. Some examples are given in Section 5. I am indebted to R. R. Bahadur, L. J. Savage, and G. Debreu for their helpful comments.

2. Problem formulation and a basic lemma. Let y be a random N-dimensional column vector, with a distribution p that is known to be in some family P of distributions. Let $m_p = Ey$ denote the mean of the distribution p, and suppose that one is required to estimate the value of $f'm_p$ on the basis of a single observation on y, where f is given. It is assumed that the loss due to incorrect estimation is the square of the error. In this note minimax estimators of $f'm_p$ will be de-

rived under two different assumptions about P; both assumptions have the following form:

Let T be given $N \times M$ matrix; let C_p denote the covariance of p, i.e.,

$$C_p = E(y - m_p)(y - m_p)';$$

and let H be a given set of $N \times N$ covariance matrices.

(2.1) For every p in P, the mean $m_p = Tx$ for some M-dimensional vector x and C_p is in H.

(2.2) For every x, and every C in H, there is a *normal* distribution in P with mean Tx and covariance C.

The assumption that P includes normal distributions is a natural one, since normality can rarely be ruled out as preposterous.

If α is any estimator, then the risk, or expected loss, associated with using α is, for any p, given by

$$(2.3) \qquad \begin{aligned} r(\alpha, p) &= E[\alpha(y) - f'm_p]^2 \\ &= E[\alpha(y) - E\alpha(y)]^2 + [E\alpha(y) - f'm_p]^2. \end{aligned}$$

An estimator $\hat{\alpha}$ is minimax if, for every estimator α,

$$\sup_{p \varepsilon P} r(\hat{\alpha}, p) \leqq \sup_{p \varepsilon P} r(\alpha, p).$$

Because of the convexity of the risk function, it is not necessary to consider randomized estimators (see [3], Theorem 3.2).

Relative to a given covariance C, an estimator α is said to be *minimum variance linear unbiased*, or more briefly, *Markoff*, if

$$(2.4) \qquad \alpha(y) = a'y \quad \text{(linearity)}.$$

$$(2.5) \qquad \text{For every } p \text{ in } P,\ Ea'y = m_p \quad \text{(unbiasedness)}.$$

(2.6) If β is any estimator satisfying (2.4) and (2.5), then for every p in P with covariance C, $r(\alpha, p) \leqq r(\beta, p)$.

The significance of the Markoff estimators in this problem is that, in both cases considered in this note, there is a Markoff estimator, relative to some C in H, that is minimax.

It follows from (2.1) that a linear estimator a is unbiased if and only if $T'a = T'f$; and from (2.3) that the risk for a linear unbiased estimator is $a'C_pa$. Therefore, a linear estimator a is Markoff relative to C if and only if it minimizes $a'Ca$ subject to the constraint $T'a = T'f$.

It might be noted here that it follows from (2.3) that the standard definition of a Markoff estimator given above is equivalent to another one in which condition (2.5) (unbiasedness) is replaced by the following (bounded risk):

(2.5′) The risk $E(a'y - f'm_p)^2$ is bounded as p varies in the class of all p in P that have covariance C, for any given C.

The idea of replacing the constraint of unbiasedness by the constraint of bounded risk is close to the minimax spirit, and seems to be due to L. J. Savage.

The main tool that will be used is the following lemma, which is closely related to a theorem of Hodges and Lehman ([3], theorem 6 5), and is stated here without proof.

LEMMA. *If $\hat{a}$ is Markoff relative to $\hat{C}$ in H, and if $\hat{a}'C\hat{a} \leqq \hat{a}'\hat{C}\hat{a}$ for every C in H, then $\hat{a}$ is minimax.*

In the "classical" situation to which the general Markoff theorem on least squares is applied (see, for example, Aitken [1]), it is assumed that the covariance of the distribution p is known up to multiplication by a positive constant, i.e., that the covariance is cC, where C is known but c is not. If it is further assumed that c is bounded by some number k, then it follows immediately from the Lemma that the Markoff estimator relative to kC is minimax Note that the Markoff estimator is independent of k.

On the other hand, if nothing at all is known about the covariance of p, i.e., if H is taken to be the class of *all* $N \times N$ covariance matrices, then the risk for every estimator is unbounded. To get a finite minimax value, the class H must be "bounded" in some sense, and the next two sections explore two directions in which such a bound can be defined In each case it should be borne in mind that postulated assumptions are thought of as applying *after, possibly, an appropriate transformation of the coordinate system.*

3. The case of bounds in terms of characteristic roots. In this section minimax estimators are derived for the problem formulated in Section 2, when the covariances are bounded in certain ways in terms of their characteristic roots.

For any covariance matrix C, let r_i denote its characteristic roots (these will be non-negative real numbers). For any number $q \geqq 1$, the *q-norm* of C is defined here to be

$$N(C; q) = \left(\sum_i r_i^q\right)^{1/q}.$$

For $q = 1, 2$, and ∞, one gets the trace of C, the square root of the sum of squares of the elements of C, and the largest characteristic root of C, respectively. Note that for the identity matrix I, $N(I; q) = N^{1/q}$.

THEOREM 1. *Let q and k be given such that $1 \leqq q \leqq \infty$ and $k > 0$, and let H be the set of all covariances C such that $N(C; q) \leqq k$; then for the estimation problem described in Section 2, the Markoff estimator $\hat{a}$ relative to the identity matrix[2] is minimax, and the minimax loss is $k\hat{a}'\hat{a}$.*

PROOF. The idea of the proof is to show that the covariance of rank one ...

[2] Strictly speaking, relative to the identity times an appropriate constant, since the identity may not be in H.

concentrates all the variance in the direction of $f'y$ is least favorable. Let $B = \hat{a}\hat{a}'/\hat{a}'\hat{a}$. Note that $N(B; q) = 1$. Since $\hat{a}$ is that unbiased linear estimator with minimum length, any unbiased linear estimator is of the form $\hat{a} + d$, where $\hat{a}'d = 0$. Hence for all unbiased linear estimators b,

$$b'Bb = \hat{a}'B\hat{a} = \hat{a}'\hat{a};$$

in particular, $\hat{a}$ is Markoff with respect to B, and to kB.

Let C be any covariance in H, and let r be its largest characteristic root; then

$$\hat{a}'C\hat{a} \leqq r\hat{a}'\hat{a} = N(C; \infty)\hat{a}'\hat{a} \leqq N(C; q)\hat{a}'\hat{a} \leqq k\hat{a}'B\hat{a}. \tag{3.1}$$

The theorem now follows from the lemma, equation (3.1), and the fact that $\hat{a}$ is Markoff relative to kB.

For the case $q = 1$, it can be shown that the minimax estimator is not unique, but it is not known whether it is unique for $q > 1$. However, the Markoff estimator $\hat{a}$ of Theorem 1 is the only linear minimax estimator, which can be seen as follows. A linear minimax estimator d must have bounded risk, and therefore must be unbiased. Suppose d is different from $\hat{a}$, and let $D = dd'/d'd$; then

$$kd'Dd = kd'd > ka'a,$$

i.e., the risk for d against the covariance kD is greater than the minimax risk.

Note that it follows immediately from Theorem 1, that if the characteristic roots of the covariance matrices in H are defined relative to any fixed symmetric positive definite matrix Q, then the Markoff estimator relative to Q will be minimax.

4. The case of bounds on the variances of the coordinates. In this section minimax estimators are found for the problem of Section 2 in the case in which the class H of covariances is delimited by bounding the variances of given linear functions of the random vector, in other words, by choosing a particular coordinate system and bounding the variances of the colrdinates.

THEOREM 2. Let $k_1, \cdots, k_N$ *be* N *given positive numbers; let* H *be the set of covariances* C *such that* $c_{ii} \leqq k_i^2$ *for* $i = 1, \cdots, N$; *then any* $\hat{a}$ *that minimizes* $\sum_i k_i \mid a_i \mid$ *subject to* $T_a' = T'f$ *(unbiasedness) is a minimax estimator for the problem of Section* 2, *and* $\hat{c}^2 = \left(\sum_i k_i \mid a_i \mid\right)^2$ *is the minimax loss.*

PROOF. There is no loss of generality in assuming that $k_i = 1$ for every i. As in Theorem 1, one is led to look for a least favorable covariance matrix among those of rank 1.

Let U be the set of linear unbiased estimators; for any C in H and b in U,

$$b'Cb = \sum_{ij} b_i b_j c_{ij} \leqq \sum_{ij} \mid b_i b_j \mid (c_{ii} c_{jj})^{\frac{1}{2}} \leqq \sum_{ij} \mid b_i b_j \mid = \left(\sum_i \mid b_i \mid\right)^2. \tag{4.1}$$

Let $\hat{a}$ be any vector that minimizes $\sum_i \mid a_i \mid$ in U, and let $\hat{c} = \sum_i \mid \hat{a}_i \mid$. By equation (4.1), and the lemma, the present theorem is proved if a vector $\hat{e}$ can be found such that (1) $\hat{a}$ is Markoff against $\hat{E} = \hat{e}\hat{e}'$; (2) $\hat{E}$ is in H, i.e., $\hat{e}_i^2 = 1$ for every i; and (3) the risk for $\hat{a}$ against E equals $\hat{c}^2$.

To this end, let S be the set of all vectors b such that $\sum_i \mid b_i \mid \leqq c$. S is a bounded convex polyhedron, and the intersection of S with U is contained in

the boundary of S, by the definition of c. Hence there is a hyperplane supporting S that contains U, i.e., there is a vector $\hat{e}$ such that

$$b'\hat{e} = \hat{c}, \text{ for all } b \text{ in } U, \tag{4.2}$$

$$b'\hat{e} \leqq \hat{c}, \text{ for all } b \text{ in } S \tag{4.3}$$

(see, for example, [2], p. 4).

By (4.2), e satisfies conditions (1) and (3) above. By the definition of S, any vector with one coordinate equal in absolute value to $\hat{c}$, and all other coordinates zero, is in S. Hence, by (4.3), $\hat{c} \mid \hat{e}_i \mid \leqq \hat{c}$, for every i, so that $\hat{e}_i^2 \leqq 1$ for every i; thus condition (2) above is also satisfied, which completes the proof.

Note that Theorem 2 characterizes all the linear minimax estimators, which is easily seen by an argument similar to that which follows Theorem 1.

5. Examples.

1. Suppose that the random variables $y_1, \cdots, y_N$ each have the same mean x, which is to be estimated, and assume that the sum of the variances of the y_i is not greater than k. To apply Theorem 1, Take T to be the $N \times 1$ matrix whose elements are all equal to 1, f to be vector for which $\sum f_i = 1$ (e.g., $[1, 0, \cdots, 0]$), and $q = 1$. It follows that a minimax estimate of $f'm_p = x$ is the arithmetic mean of $y_1, \cdots, y_N$, i.e., $\hat{a} = (1/N, \cdots, 1/N)$, and the minimax loss is $k \sum_i \hat{a}_i^2 = k/N$ This minimax estimator is, of course, the Markoff estimator for the situation in which it is known that the y_i are independent, with equal variances.

The same result would be obtained if it were assumed that the variance of any linear combination $\sum b_i y_i$ such that $\sum b_i^2 = 1$ is not greater than k (the case $q = \infty$).

2. Consider the estimation problem of Example 1, except now assume that the variance of y_i is not greater than k_i^2, $i = 1, \cdots, N$. By Theorem 2, a minimax estimator is given by

$$\hat{a}_i = \begin{cases} 1, & \text{for that } i \text{ for which } k_i \text{ is minimum,} \\ 0, & \text{otherwise,} \end{cases} \tag{5.1}$$

and the minimax loss is $\min_i k_i^2$. Note that in this example the minimax loss is independent of the sample size N, except insofar as $\min_i k_i$ depends upon N. If $k_1 = \cdots = k_N$, then any linear unbiased estimator is minimax.

3. Suppose it is required to estimate the slope e in the linear regression of one variable on another, and it is assumed that the variance of the "dependent variable" is not greater than k^2. To apply Theorem 2, take

$$T' = \begin{bmatrix} 1, \cdots, 1 \\ t_1, \cdots, t_N \end{bmatrix} \quad \text{and} \quad x' = (d, e),$$

where $t_1, \cdots, t_N$ are the values of the "independent variable," and d and e are unknown. A bounded risk (unbiased) linear estimator a must satisfy

$$\begin{aligned} \sum a_i &= 0, \\ \sum a_i t_i &= 1. \end{aligned} \tag{5.2}$$

By Theorem 2, any $\hat{a}$ that minimizes $\sum |a_i|$ subject to equation (5.2) is a minimax estimator of e. Without loss of generality, t_N can be taken to be the largest value of t_i, and t_1 the smallest; then it is not hard to show that the unique solution of the above minimization problem is

$$a_i = \begin{cases} \dfrac{-1}{t_N - t_1}, & \text{for} \quad i = 1, \\ \dfrac{1}{t_N - t_1}, & \text{for} \quad i = N, \\ 0 \quad, & \text{otherwise;} \end{cases} \tag{5.3}$$

and the minimax loss is $4k^2/(t_N - t_1)^2$. In other words, a minimax estimate of e is obtained by taking the slope of the line passing through the "extreme" points (y_1, t_1) and (y_N, t_N).

4. Consider the estimation problem of Example 3, but assume that the sum of the variances of $y_1, \cdots, y_N$ is not greater than k. As in Example 1, this corresponds to taking $q = 1$ in Theorem 1. By Theorem 1 the usual least squares estimate $\sum [(y_i - \bar{y})(t_i - \bar{t})]/(t_i - \bar{t})^2$ is a minimax estimate of e, and the minimax loss is $k/\sum (t_i - \bar{t})^2$.

Suppose further that $t_i = i - 1$ (e.g., think of t_i as successive times), and consider the transformation (taking successive differences)

$$z_i = \begin{cases} y_1, & \text{for } i = 1, \\ y_i - y_{i-1}, & \text{for } 1 = 2, \cdots, N. \end{cases} \tag{5.4}$$

The means of the z_i are

$$Ez_i = \begin{cases} d, & \text{for } i = 1, \\ e, & \text{for } i = 2, \cdots, N. \end{cases} \tag{5.5}$$

Now assume that the sum of the variances of the *new variables* z_i is not greater than k; then by Theorem 1 a minimax estimate of e is

$$\frac{1}{N-1} \sum_2^N z_i = \frac{y_N - y_1}{N - 1},$$

and the minimax loss is $k/(N - 1)$, a different result from that obtained before making the transformation (5.4).

REFERENCES

[1] A. C. Aitken, "On least squares and the linear combination of observations," *Proc. Roy. Soc. Edinburgh*, Vol. 55 (1935).

[2] T. Bonnesen and W. Fenchel, *Theorie der Konvexen Körper*, J. Springer, Berlin, 1934.

[3] J. L. Hodges and E. L. Lehmann, "Some problems in minimax point estimation," *Ann. Math. Stat.*, Vol. 21 (1950), 182–197.

[4] L. J. Savage, *The Foundations of Statistics*, John Wiley & Sons, New York, 1954.

COVARIANCES OF LEAST-SQUARES ESTIMATES WHEN RESIDUALS ARE CORRELATED[1]

By M. M. Siddiqui[2]

University of North Carolina

1. Summary. In this paper we will study the effects on the covariance matrix of the least-squares estimates of regression coefficients and on the estimate of the residual variance when the usual condition of independence of residuals is violated. The cases of linear trend and of regression on trigonometric functions will be considered in some detail

2. Introduction. Several authors have studied the problem of estimating regression coefficients when residuals are autocorrelated We refer here only to the work of Grenander and Rosenblatt [2, 3, 4]. Grenander [2] gives conditions on the regression variables for the existence of consistent estimates of the regression coefficients. He also gives conditions on the residual process under which the least-squares (L.S.) estimate of a regression coefficient is asymptotically efficient with respect to the Markov estimate The covariances of the L S. estimates as summarized in a matrix form are well known and are given at the end of section 3. The exact expression for an individual covariance or variance in the general case is easily extracted from this matrix and is given in section 4 The variance of the L S. estimate in the general case is also given by Grenander [2, (8) p. 258]. Asymptotic expressions for the covariances of these estimates are also available [2, 4]. However, it seemed desirable to present here, in some detail, exact expressions or high order approximations to them for the individual variances and covariances of the L.S. estimates of regression coefficients and for the expectation of the estimate of residual variance, particularly for the cases of general interest, in readily usable form, and derived in an elementary fashion. The first term of each of our expressions coincide with the asymptotic expression given in [2, 4], when the regression coefficients are made comparable.

Bounds on the covariances of L.S. estimates are also provided in (7).

3. The L.S. estimates. Let $y = x'\beta + \Delta$ be the regression equation, where y and Δ are $N \times 1$ column vectors, β is a $p \times 1$ column vector, x is a $p \times N$ matrix and a prime is used to denote the transpose of a matrix or a vector. It is assumed that $N > p$, x is non-stochastic and of rank p, and Δ is a $N(0, \sigma^2 P)$ vector variate, where 0 is a zero vector and P is a positive definite correlation matrix.

Introducing $c = (xx')^{-1}$, $b = cxy$, $v = y - x'b$, $n = N - p$, and writing $\hat{\sigma}$

Received December 3, 1957; revised June 13, 1958.

[1] Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at the Boulder Laboratories, National Bureau of Standards.

for $q - Eq$, where q is a variate with expected value Eq, it is known that b and $s^2 = v'v/n$ are the least-squares estimates of β and σ^2 respectively. It is also known that $Eb = \beta$, and

$$B = E\delta b\delta b' = \sigma^2 cxPx'c.$$

In case $P = I_N$, where I_N is the $N \times N$ identity matrix, $Es^2 = \sigma^2$ and $B = \sigma^2 c$.

4. The covariance matrix B. We propose to study the effects on B and Es^2 when P is given by

(1) $$P = I_N + \sum_{k=1}^{N-1} \rho_k (C^k + C'^k),$$

where

(2) $$C = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

i.e. when

(3) $$E\Delta_t \Delta_{t\pm k} = \sigma^2 \rho_k, \qquad k = 0, 1, \cdots, N-1, \rho_0 = 1.$$

We have

(4) $$v = y - x'b = x'(\beta - b) + \Delta = (I_N - x'cx)\Delta$$

as $b = \beta + cx\Delta$. Writing $m = x'cx$, we have $m' = m$ and $m^2 = m$. Hence if λ is a characteristic root of m, $\lambda = 0$ or 1. Writing "tr" for the trace of a matrix we obtain tr $m = p$. Now, by simple evaluation

(5) $$Es^2 = \frac{\sigma^2}{n}[N - \operatorname{tr} Pm] = \sigma^2 \left[1 - \frac{2}{n} \sum_{k=1}^{N-1} \sum_{t=1}^{N-k} \rho_k m_{t,t+k} \right].$$

Here, if e is a matrix, e_{ij} or $e_{i,j}$ refers to its element in the ith row and the jth column.

If we write $d = cx$, we find that

$$B_{ij} = E\delta b_i \delta b_j = \sigma^2 \left[\sum_{t=1}^{N} d_{it} d_{jt} + \sum_{k=1}^{N-1} \sum_{t=1}^{N-k} \rho_k (d_{jt} d_{i,t+k} + d_{it} d_{j,t+k}) \right].$$

If, by a proper choice of x or with a suitable transformation on x, we make $xx' = c^{-1} = I_p$, we have $d = x$. Writing x_i' for the row vector in the ith row of x, we find

(6) $$B_{ij} - \sigma^2 \delta_{ij} = \sigma^2 \sum_{k=1}^{N-1} \rho_k x_i'(C^k + C'^k) x_j, \qquad i, j = 1, \cdots, p;$$

where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$.

It has been shown [1, p. 130] that if A is an $N \times N$ real symmetric matrix with

characteristic roots $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_N$, and u and v are $N \times 1$ real vectors, then, under the conditions $u'u = v'v = 1$, $u'v = 0$, the bilinear form $u'Ay$ has a maximum $(\alpha_N - \alpha_1)/2$ and a minimum $(\alpha_1 - \alpha_N)/2$. Also the quadratic form $u'Au \leqq \alpha_N$. Now the maximum characteristic root of $C^k + C'^k$, where k is a positive integer,

$$\alpha_N = 2\cos\left\{\frac{\pi}{\left[\frac{N+k-1}{k}\right]+1}\right\} \leqq 2\cos\left\{\frac{k\pi}{N+2k-1}\right\},$$

where $[q]$ denotes the largest integer $\leqq q$, and the minimum characteristic root $\alpha_1 = -\alpha_N$, [1, p. 101]. Hence, we obtain

$$\text{(7)} \qquad |B_{ij} - \sigma^2\delta_{ij}| < 2\sigma^2 \sum_{k=1}^{N-1} \left| \rho_k \cos\frac{k\pi}{N+2k-1} \right| < 2\sigma^2 \sum_{k=1}^{N-1} |\rho_k|.$$

In the case $\rho_k = \rho^k$, $k = 0, 1, \cdots$, where $\alpha = |\rho| < 1$, we have

$$\text{(8)} \qquad |B_{ij}| < \frac{2\alpha\sigma^2}{1-\alpha} \quad \text{if } i \neq j, \qquad B_{ii} < \sigma^2\left(\frac{1+\alpha}{1-\alpha}\right).$$

5. Linear trend. Let $N = 2r + 1$ where r is a positive integer and consider the linear trend in the form

$$\text{(9)} \qquad y_t = \beta_1(2r+1)^{-1/2} + \beta_2(t-r-1)/a + \Delta_t, \qquad t = 1, \cdots, N,$$

where

$$\text{(10)} \qquad a^2 = r(r+1)(2r+1)/3 = N(N^2-1)/12.$$

In the notation of section 3

$$\text{(11)} \qquad \begin{aligned} x_{1t} &= (2r+1)^{-1/2}, \qquad x_{2t} = (t-r-1)/a, \qquad t = 1, \cdots, N, \\ b_1 &= \sqrt{N}\bar{y}, \qquad b_2 = \left[\sum_{t=1}^{N} ty_t - (r+1)\sum_{t=1}^{N} y_t\right] \Big/ a. \end{aligned}$$

Furthermore

$$c = I_p, \qquad p = 2, \qquad n = N - 2,$$

$$m_{ij} = (x'x)_{ij} = \frac{1}{N} + \frac{3(2i-N-1)(2j-N-1)}{N(N^2-1)}$$

$$ns^2 = \sum y_t^2 - b_1^2 - b_2^2,$$

$$\text{(12)} \qquad B_{11} = \sigma^2\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_k\right], \qquad B_{12} = 0,$$

$$B_{22} = \sigma^2\left[1 + 2\sum_{k=1}^{N-1}\rho_k - \frac{2}{N}\left(3 + \frac{2}{N^2-1}\right)\sum_{k=1}^{N-1}k\rho_k + \frac{4}{N(N^2-1)}\sum_{k=1}^{N-1}k^3\rho_k\right]$$

and

$$Es^2 = \sigma^2\left[1 - \frac{4}{n}\sum_{k=1}^{N-1} \rho_k + \frac{2}{nN}\left(4 + \frac{2}{N^2-1}\right)\sum_{k=1}^{N-1} k\rho_k - \frac{4}{nN(N^2-1)}\sum_{k=1}^{N-1} k^3\rho_k\right]$$

For the case when $\rho_k = \rho^k$, we can evaluate the summations $\sum \rho_k$, $\sum k\rho_k$, etc. If N is moderately large we may neglect ρ^N and thus find

$$(13)\quad \begin{aligned} E\frac{s^2}{\sigma^2} &\cong 1 - \frac{4\rho}{n(1-\rho)} + \frac{8\rho}{nN(1-\rho)^2} + \frac{4(\rho + 4\rho^2 + \rho^3)}{nN(N^2-1)(1-\rho)^4}, \\ \frac{B_{11}}{\sigma^2} &= 1 - \frac{2}{N}\frac{\rho - N\rho^N + (N-1)\rho^{N+1}}{(1-\rho)^2} + \frac{2(\rho - \rho^N)}{1-\rho} \cong \frac{1+\rho}{1-\rho} - \frac{2\rho}{N(1-\rho)^2}, \\ B_{12} &= 0, \qquad \frac{B_{22}}{\sigma^2} \cong \frac{1+\rho}{1-\rho} - \frac{6\rho}{N(1-\rho)^2}. \end{aligned}$$

We note that b_i are independently distributed $N(\beta_i, B_{ii})$, $i = 1, 2$, variates. If we set

$$b_1' = N^{-1/2}b_1 = \bar{y}, \qquad b_2' = \frac{\sqrt{12}b_2}{\sqrt{N(N^2-1)}},$$

the estimate of Ey_t is given by

$$(14)\qquad Y_t = \bar{y} + b_2'(t - r - 1)$$

and under the first order autoregressive scheme for Δ's,

$$(15)\qquad \sigma_{\bar{y}}^2 \cong \frac{\sigma^2}{N}\left(\frac{1+\rho}{1-\rho}\right), \qquad \sigma_{b_2'}^2 \cong \frac{12\sigma^2}{N(N^2-1)}\left(\frac{1+\rho}{1-\rho}\right), \qquad \operatorname{cov}(\bar{y}, b_2') = 0.$$

Thus

$$\sigma_{Y_t}^2 \cong \frac{\sigma^2}{N}\left(\frac{1+\rho}{1-\rho}\right)\left[1 + \frac{12(t-r-1)^2}{N^2-1}\right].$$

6. Regression on trigonometric functions. Consider

$$(16)\quad \begin{aligned} y_t = \beta_1/\sqrt{N} &+ \sqrt{2/N}\sum_{i=1}^{q} \beta_{2i}\cos\lambda_i t \\ &+ \sqrt{2/N}\sum_{i=1}^{q} \beta_{2i+1}\sin\lambda_i t + \Delta_t, \qquad t = 1, \cdots, N, \end{aligned}$$

where $\lambda_i = 2\pi\omega_i/N$ and ω_i is a positive integer less than N for $i = 1, 2, \cdots, q$ and $\omega_i \neq \omega_j$ if $i \neq j$.

In the notation of section 2

$$x_{1t} = 1/\sqrt{N}, \qquad x_{2i,t} = \sqrt{2/N} \cos \lambda_i t,$$
$$x_{2i+1,t} = \sqrt{2/N} \sin \lambda_i t, i = 1, 2, \cdots, q; t = 1, 2, \cdots, N,$$
$$xx' = c^{-1} = I_{2q+1}, \qquad n = N - 2q - 1,$$
$$b_1 = \sqrt{N}\bar{y}, \qquad b_{2i} = \sqrt{2/N} \sum_t y_t \cos \lambda_i t,$$
$$(17) \quad b_{2i+1} = \sqrt{2/N} \sum_t y_t \sin \lambda_i t, i = 1, \cdots, q,$$
$$m_{ts} = 1/N + 2/N \sum_{i=1}^{q} \cos (t - s)\lambda_i, \qquad t, s = 1, \cdots, N,$$
$$s^2 = \left(\sum_t y_t^2 - \sum_{i=1}^{2q+1} b_i^2\right) \Big/ n.$$

We find

$$(18) \quad E\frac{s^2}{\sigma^2} = 1 - \frac{2}{n}\sum_{k=1}^{N-1} \rho_k + \frac{2}{nN}\sum_{k=1}^{N-1} k\rho_k - \frac{4}{n}\sum_{k=1}^{N-1}\sum_{i=1}^{q}\left(1 - \frac{k}{N}\right)\rho_k \cos k\lambda_i.$$

For the covariances of b_i and b_j we obtain

$$B_{11} = \sigma^2\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_k\right],$$
$$B_{1,2i} = \frac{\sqrt{2}\,\sigma^2}{N}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k} \rho_k \{\cos (t + k)\lambda_i + \cos t\lambda_i\}, \qquad i = 1, \cdots, q,$$
$$(19) \quad B_{2i,2i} = \sigma^2\left[1 + \frac{4}{N}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k} \rho_k \cos (k + t)\lambda_i \cos t\lambda_i\right], \qquad i = 1, \cdots, q.$$
$$B_{2i,2j+1} = \frac{2\sigma^2}{N}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k} \rho_k \{\cos t\lambda_i \sin (k + t)\lambda_j$$
$$+ \cos (k + t)\lambda_i \sin t\lambda_j\}, \qquad i,j = 1, \cdots, q.$$

$B_{1,2i+1}$ and $B_{2i+1,2i+1}$ are obtainable from the expressions for $B_{1,2i}$ and $B_{2i,2i}$ respectively by replacing cosine by sine.

If $\rho_k = \rho^k$, and ρ^N is negligible, we find for the variances, after some reduction,

$$\frac{B_{11}}{\sigma^2} \cong \frac{1 + \rho}{1 - \rho} - \frac{2\rho}{N(1 - \rho)^2},$$
$$(20) \quad \frac{B_{2i,2i}}{\sigma^2} \cong \frac{1 - \rho^2}{1 - 2\rho\cos\lambda_i + \rho^2} + \frac{\rho\cos\lambda_i}{N(1 - 2\rho\cos\lambda_i + \rho^2)} - \frac{\rho(1 + \rho^2)\cos\lambda_i - 2\rho^2}{N(1 - 2\rho\cos\lambda_i + \rho^2)^2},$$
$$\frac{B_{2i+1,2i+1}}{\sigma^2} \cong \frac{1 - \rho^2}{1 - 2\rho\cos\lambda_i + \rho^2} - \frac{\rho\cos\lambda_i}{N(1 - 2\rho\cos\lambda_i + \rho^2)} - \frac{\rho(1 + \rho^2)\cos\lambda_i - 2\rho^2}{N(1 - 2\rho\cos\lambda_i + \rho^2)^2}, \qquad i = 1, \cdots, q.$$

Also

$$(21) \qquad E\frac{s^2}{\sigma^2} \cong 1 - \frac{2\rho}{n(1-\rho)} - \frac{4\rho}{n}\sum_{i=1}^{q}\frac{\cos\lambda_i - \rho}{1 - 2\rho\cos\lambda_i + \rho^2} + O\left(\frac{1}{N^2}\right).$$

7. Concluding remarks. We conclude with the remarks that in most practica cases the correlation matrix for Δ's will not be known. However, if Δ's may b represented as a stationary autoregressive process of some small order—in man cases first or second order scheme gives a reasonably good fit—we would be re quired to estimate a few parameters $\rho_1, \rho_2, \cdots, \rho_k$. We, however, note tha these quantities do not appear in b and s^2, only in B and Es^2.

We further observe that the estimates, $\hat{\beta}$ and $\hat{\sigma}^2$, of β and σ^2 obtained fron maximizing the likelihood function will depend on the parameters of P, i.e. o $\rho_1, \rho_2, \cdots, \rho_{N-1}$, which will mean using sample serial correlation coefficient to estimate ρ's in the expression for $\hat{\beta}$ and $\hat{\sigma}^2$. These estimates will obviously b non-linear. Thus it seems more desirable to stick to the least-squares estimate b and s^2 rather than to attempt to develop maximum-likelihood (or minimun χ^2) estimates.

8. Acknowledgement. The writer wishes to express his indebtedness t Professor Harold Hotelling for drawing his attention to this problem.

REFERENCES

[1] M. M. SIDDIQUI, *Distributions of Some Serial Correlation Coefficients*, University of Nort Carolina Institute of Mathematical Statistics Mimeograph Series No. 164, Chape Hill (1957).

[2] ULF GRENANDER, "On the estimation of regression coefficients in the case of an auto correlated disturbance," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 252–272.

[3] MURRAY ROSENBLATT, "On the estimation of regression coefficients of a vector-value time series with a stationary residual," *Ann. Math. Stat.*, Vol. 27 (1956), pp 99–121.

[4] ULF GRENANDER AND MURRAY ROSENBLATT, *Statistical Analysis of Stationary Tim Series*, John Wiley and Sons, New York, 1957, Chapter 7.

ON A PROBABILITY PROBLEM IN THE THEORY OF COUNTERS

By L. Takács

Research Institute for Mathematics, Hungarian Academy of Sciences, Budapest, Hungary

1. Introduction. Let us suppose that particles arrive at a counter in the time interval $(0, \infty)$ according to a Poisson-process of density λ. Each particle arriving in the time interval $(0, \infty)$ independently of the others gives rise to an impulse with probability p or 1 according to whether at this instant there is an impulse present or there is no impulse present. The time durations of the impulses are identically distributed independent positive random variables with distribution function $H(x)$ and these random variables are independent of the instants of the arrivals and of the events of the realizations of the impulses. We define as "registered particles" those particles which occur at an instant when there is no impulse present. Denote by ν_t the number of the registered particles in the time interval $(0, t)$. The problem is to determine the distribution law of ν_t and its asymptotic behaviour as $t \to \infty$.

The particular case of the above problem, when the time durations of the impulses are constant, was investigated earlier by G. E. Albert and L. Nelson [1].

2. The structure of the process. Denote by $\{\tau_n\}$ the sequence of instants at which particles are registered. We say that the system at any instant t is in state A when no impulse covers the instant t and in state B otherwise. Then the system assumes the states $A, B, A, B, \cdots$ alternatingly. Let us denote by $\xi_1, \eta_1, \xi_2, \eta_2, \cdots$ the times spent in states A and B respectively. If the system at the instant t is in state A, then t is evidently a regeneration point of the process. Consequently $\{\xi_n\}$ and $\{\eta_n\}$ are independent sequences of identically distributed positive random variables. Clearly $\mathbf{P}\{\xi_n \leqq x\} = F(x) = 1 - e^{-\lambda x}$ if $x \geqq 0$. Write $\mathbf{P}\{\eta_n \leqq x\} = U(x)$, where $U(x)$ is still unknown. (We use $\mathbf{P}$ for the symbol of probability and $\mathbf{E}$ for expectation.) It can easily be seen that the instants of the transitions $A \to B$ coincide with the instants τ_n $(n = 1, 2, \cdots)$. Consequently the time differences $\tau_{n+1} - \tau_n$ $(n = 1, 2, \cdots)$ are identically distributed independent random variables with distribution function $G(x) = F(x) * U(x)$ i.e.

$$G(x) = \int_0^x U(x - y)e^{-\lambda y}\lambda\, dy, \tag{1}$$

while $\mathbf{P}\{\tau_1 \leqq x\} = F(x)$.

3. Notations. Let us introduce the following Laplace-Stieltjes transforms:

$$\gamma(s) = \int_0^\infty e^{-sx}\, dG(x) \tag{2}$$

Received May 17, 1957; revised May 15, 1958.

and

(3) $$\omega(s) = \int_0^\infty e^{-sx}\, dU(x).$$

By (1) we have

(4) $$\gamma(s) = \frac{\lambda}{\lambda + s}\,\omega(s).$$

Further put

(5) $$\alpha = \int_0^\infty x\, dH(x), \qquad \beta^2 = \int_0^\infty (x - \alpha)^2\, dH(x),$$

(6) $$\tau = \int_0^\infty x\, dU(x), \qquad \rho^2 = \int_0^\infty (x - \tau)^2\, dU(x),$$

(7) $$A = \int_0^\infty x\, dG(x), \qquad B^2 = \int_0^\infty (x - A)^2\, dG(x).$$

By (1) we clearly have that $A = \tau + (1/\lambda)$ and $B^2 = \rho^2 + (1/\lambda^2)$.

Denote by $P(t)$ the probability that at the instant t the system is in state A, and put

(8) $$\pi(s) = \int_0^\infty e^{-st} P(t)\, dt.$$

4. Theorems concerning ν_t. In what follows we shall give some general theorems for ν_t.

1. We have

(9) $$\mathbf{P}\{\nu_t \leqq n\} = 1 - F(t) * G_n(t),$$

where $G_n(x)$ denotes the n-fold convolution of $G(x)$ with itself. ($G_0(x) = 1$ if $x \geqq 0$ and $G_0(x) = 0$ if $x < 0$). For

$$\mathbf{P}\{\nu_t \leqq n\} = \mathbf{P}\{t < \tau_{n+1}\} = 1 - \mathbf{P}\{\tau_{n+1} \leqq t\},$$

and $\tau_{n+1} = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_{n+1} - \tau_n)$ is a sum of independent random variables.

2. If $A < \infty$, then we have

(10) $$\lim_{T\to\infty} \mathbf{P}\{\nu_{T+t} - \nu_T \leqq n\} = 1 - G^*(t) * G_n(t),$$

where

$$G^*(t) = \begin{cases} \dfrac{1}{A}\displaystyle\int_0^t [1 - G(u)]\, du & \text{if} \quad t \geqq 0 \\ 0 & \text{if} \quad t < 0. \end{cases}$$

The proof is similar to that of (9), only we must use the result

$$\lim_{T\to\infty} \mathbf{P}\{\nu_{T+t} - \nu_T \geqq 1\} = G^*(t),$$

which was proved by J. L. Doob [2].

3. If $B^2 < \infty$, then we have

$$(11) \qquad \lim_{t\to\infty} \mathbf{P}\left\{\frac{\nu_t - \frac{t}{A}}{\sqrt{\frac{B^2 t}{A^3}}} \leqq x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du.$$

This can be proved by the aid of the method of W. Feller [3]. (Cf. [5]).

4. If $B^2 < \infty$, then we have

$$(12) \qquad \mathbf{P}\left\{\limsup_{t\to\infty} \frac{\nu_t - \frac{t}{A}}{\sqrt{\frac{2B^2}{A^3}\, t \log\log t}} = 1\right\} = \mathbf{P}\left\{\liminf_{t\to\infty} \frac{\nu_t - \frac{t}{A}}{\sqrt{\frac{2B^2}{A^3}\, t \log\log t}} = -1\right\} = 1.$$

This can be proved by the aid of the law of the iterated logarithm stated by P Hartman and A. Wintner [4].

5. Applying the strong law of large numbers we obtain

$$(13) \qquad \mathbf{P}\left\{\lim_{t\to\infty} \frac{\nu_t}{t} = \frac{1}{A}\right\} = 1,$$

(cf. J. L. Doob [2]).

It is easy to see that $\mathbf{E}\{\nu_t\} = M(t)$ can be expressed as follows:

$$(14) \qquad M(t) = \sum_{n=1}^{\infty} \mathbf{P}\{\tau_n \leqq t\}.$$

6. If $A < \infty$, then for any $h > 0$ we have

$$(15) \qquad \lim_{t\to\infty} \frac{M(t+h) - M(t)}{h} = \frac{1}{A},$$

by the theorem of J. L. Doob [2].

7. If $B^2 < \infty$, then we have

$$\int_0^\infty e^{-st}\, dM(t) = \frac{1}{As} + \frac{B^2 + A^2}{2A^2} - \frac{1}{\lambda A} + o(s) \tag{16}$$

if $s \to 0$. For by (9) and (14) we have

$$\int_0^\infty e^{-st}\, dM(t) = \frac{\lambda}{(\lambda + s)[1 - \gamma(s)]} \tag{17}$$

and

$$\gamma(s) = 1 - sA + \frac{s^2}{2}(B^2 + A^2) + o(s^2)$$

if $s \to 0$.

8. For the Laplace-transform of $P(t)$ we have

$$\pi(s) = \int_0^\infty e^{-st} P(t)\, dt = \frac{1}{(\lambda + s)[1 - \gamma(s)]}, \tag{18}$$

and

$$P = \lim_{t\to\infty} P(t) = \frac{1}{\lambda A}. \tag{19}$$

Proof. As $M(t + \Delta t) = M(t) + P(t)\lambda\Delta t + o(\Delta t)$, we have $M'(t) = \lambda P(t)$, and thus (18) follows from (17). Now

$$P(t) = 1 - \int_0^t [1 - U(t - x)]\, dM(x), \tag{20}$$

for by the theorem of total probability we have

$$1 - P(t) = \sum_{n=1}^{\infty} \int_0^t [1 - U(t - x)]\, d\mathrm{P}\{\tau_n \leqq x\} = \int_0^t [1 - U(t - x)]\, dM(x),$$

which agrees with (20). By virtue of (15) we obtain from (20)

$$\lim_{t\to\infty} P(t) = 1 - \frac{1}{A}\int_0^\infty [1 - U(x)]\, dx = 1 - \frac{\tau}{A}.$$

Since $\tau = A - (1/\lambda)$, equation (19) follows.

Remark. Taking into consideration that $M'(t) = \lambda P(t)$, we obtain from (20) the following integral equation for $P(t)$:

$$P(t) = 1 - \lambda \int_0^t [1 - U(t - x)] P(x)\, dx. \tag{21}$$

From (18) or from (21) we obtain that

$$\omega(s) = \int_0^\infty e^{-sx}\,dU(x) = \frac{\lambda + s}{\lambda}\left[1 - \frac{1}{(\lambda + s)\pi(s)}\right]. \tag{22}$$

To apply the above theorems it remains only to determine $G(x)$, A, and B^2.

5. The determination of $G(x)$, A, and B^2.

THEOREM. *If $0 < p \leqq 1$, then we have*

$$\begin{aligned}\gamma(s) = \int_0^\infty e^{-sx}\,dG(x) &= \frac{\lambda p + s}{p(\lambda + s)} \\ &- \frac{1}{p(\lambda + s)}\left\{\int_0^\infty \exp\left[-st - \lambda p \int_0^t (1 - H(x))\,dx\right]dt\right\}^{-1};\end{aligned} \tag{23}$$

if $\alpha < \infty$, then

$$A = \frac{e^{\lambda p\alpha} + p - 1}{\lambda p}, \tag{24}$$

and if $\beta^2 < \infty$, then

$$\begin{aligned}B^2 = \frac{2e^{\lambda p\alpha}}{\lambda p}\int_0^\infty &\left\{\exp\left[\lambda p \int_t^\infty (1 - H(x))\,dx\right] - 1\right\}dt \\ &+ \frac{2e^{\lambda p\alpha} - e^{2\lambda p\alpha} + p^2 - 1}{(\lambda p)^2}.\end{aligned} \tag{25}$$

If $p = 0$ then $U(x) = H(x)$ and consequently

$$G(x) = \int_0^x H(x - y)e^{-\lambda y}\lambda\,dy, \tag{26}$$

$$A = \frac{1 + \lambda\alpha}{\lambda}, \tag{27}$$

and

$$B^2 = \frac{1 + \lambda^2\beta^2}{\lambda^2}. \tag{28}$$

PROOF. Let us consider a new process which is a particular case of the process defined in the Introduction. Suppose that the density of the underlying Poisson-process is λ^* and each particle gives rise to an impulse (with probability $p^* = 1$). Let $H^*(x) = H(x)$ be the distribution function of the duration of the impulses. This is the case of Type II counter. Denote by $\{\xi_n^*\}$ and $\{\eta_n^*\}$ the sequences of the times spent in state A and B respectively. Clearly $\mathbf{P}\{\xi_n^* \leqq x\} = 1 - e^{-\lambda^* x}$ if $x \geqq 0$. Write $\mathbf{P}\{\eta_n^* \leqq x\} = U^*(x)$. Denote by $P^*(t)$ the probability that at the instant t there is no impulse present. We have showed in [5] that

$$P^*(t) = \exp\left[-\lambda^* \int_0^t [1 - H(x)]\,dx\right]. \tag{29}$$

Applying (22) it follows that

$$\omega^*(s) = \int_0^\infty e^{-sx}\, dU^*(x) = \frac{\lambda^* + s}{\lambda^*} - \frac{1}{\lambda^*\pi^*(s)}, \tag{30}$$

where

$$\pi^*(s) = \int_0^\infty e^{-st}P^*(t)\, dt = \int_0^\infty \exp\left\{-st - \lambda^* \int_0^t [1 - H(x)]\, dx\right\} dt. \tag{31}$$

Now we observe that, if in this new process $\lambda^* = \lambda p$, then we have

$$U^*(x) = U(x), \tag{32}$$

where $U(x)$ is related to the general process. The equality (32) can easily be seen if we take into consideration that the arrivals of those particles in the general process, which arrive during a dead time and which give rise to an impulse, form a Poisson process with density λp. Accordingly by (30) and (31) we have

$$\begin{aligned}\omega(s) = \int_0^\infty e^{-sx}\, dU(x) &= \frac{\lambda p + s}{\lambda p} \\ &- \left\{\lambda p \int_0^\infty \exp\left[- st - \lambda p \int_0^t (1 - H(x))\, dx\right] dt\right\}^{-1},\end{aligned} \tag{33}$$

and by (4) we obtain (23), which was to be proved.

If we introduce for the new process the analogous quantities $M^*(t)$, A^* and B^{*2} corresponding to (7) and (14), then by (16) we obtain that

$$\int_0^\infty e^{-st}\, dM^*(t) = \lambda^* \int_0^\infty e^{-st}\, P^*(t)\, dt = \frac{1}{A^*s} + \frac{B^{*2} + A^{*2}}{2A^{*2}} - \frac{1}{\lambda^*A^*} + o(s) \tag{34}$$

if $s \to 0$. Since $P^* = \lim_{t\to\infty} P^*(t) = e^{-\lambda^*\alpha}$, we obtain from (34) that

$$A^* = e^{\lambda^*\alpha} / \lambda^* \tag{35}$$

and, further,

$$B^{*2} = 2\lambda^*A^{*2} \int_0^\infty [P^*(t) - P^*]\, dt - A^{*2} + 2A^*/\lambda^*. \tag{36}$$

If in particular $\lambda^* = \lambda p$, then clearly

$$A - \frac{1}{\lambda} = A^* - \frac{1}{\lambda^*}$$

and

$$B^2 - \frac{1}{\lambda^2} = B^{*2} - \frac{1}{\lambda^{*2}},$$

and thus (24) and (26) are proved. The case $p = 0$ is evident.

Finally we remark that the more general case when the arrivals of the par-

ticles to the counter form a recurrent process was dealt by the author [6], [7], [8], but explicit solution was given only for a particular distribution $H(x)$.

REFERENCES

[1] G. E. ALBERT AND L NELSON, "Contributions to the statistical theory of counters data, *Ann. Math Stat.*, 24 (1953), 9–22
[2] J. L. DOOB, "Renewal theory from the point of view of the theory of probability," *Trans. Amer Math. Soc*, 63 (1948), pp 422–438
[3] W. FELLER, "Fluctuation theory of recurrent events," *Trans Amer Math Soc*, 67 (1949), pp. 98–119
[4] P. HARTMAN AND W. WINTNER, "On the law of the iterated logarithm," *Amer J Math*, 63 (1941), pp. 169–176.
[5] L. TAKÁCS, "On a probability problem arising in the theory of counters," *Proc Cambr Phil. Soc.*, 52 (1956), pp. 488–498
[6] L. TAKÁCS, "On stochastic processes in the theory of counters," (Hungarian), *Magy Tud. Akad.* III *Oszt Kozl.*, 6 (1956), pp 369–421
[7] L. TAKÁCS, "On the sequence of events, selected by a counter from a recurrent process of events," *The Theory of Probability and Its Applications* (Moscow) 1 (1956), pp. 90–102.
[8] L. TAKÁCS, "On some probability problems concerning the theory of counters," *Acta Math Acad. Sci Hung.*, 8 (1957), pp. 127–138

NOTES

DISTRIBUTION OF LINEAR CONTRASTS OF ORDER STATISTICS[1]

By Jacques St-Pierre

University of Montreal

Introduction. Many theoretical and practical problems of statistical nature have lead investigators to study methods capable of pooling the information contained in the ordered (or ranked) sample values with some properties of the assumed distribution of the parent population. Since, in analysis of variance situations, contrasts between functions of observations are of utmost importance, linear contrasts of order statistics will be considered here under the assumption that the underlying distribution is normal.

Null distribution of linear contrasts of order statistics. Let $x_0, x_1, \cdots, x_n$ denote $n + 1$ independent normal random variables with unknown means $\mu_1, \mu_2, \cdots, \mu_n$ respectively, and with a common variance $\sigma^2 = 1$ (say). Let $x_{(0)} > x_{(1)} > \cdots > x_{(n)}$ be the ordered values. Consider the following linear contrast

$$z = x_{(0)} - c_1 x_{(1)} - c_2 x_{(2)} - \cdots - c_n x_{(n)}, \qquad \sum_{i=1}^{n} c_i = 1;$$

$$0 \leqq c_i \leqq 1, \quad i = 1, \cdots, n.$$

Using, as a starting point, the joint density of $x_{(0)}, x_{(1)}, \cdots, x_{(n)}$ as given by Wilks [7], and with the help of appropriate transformations, the null distribution of z can be obtained. It takes the form of a rather messy expression containing a n-fold iterated integral. An interesting particular case: the density of the difference between the two largest ordered values can be obtained from the general form. St-Pierre and Zinger [6] have tabulated the null density of $u = x_{(0)} - x_{(1)}$ using a slightly different method.

It is of interest to consider the above contrast in the case of three random variables. The density of $z = x_{(0)} - cx_{(1)} - (1 - c)x_{(2)}$, under the hypothesis $H_0: \mu_0 = \mu_1 = \mu_2 = 0$ (say), takes the form

$$g(z) = 3[\pi(c^2 - c + 1)]^{-1/2} \exp\,[-z^2/4(c^2 - c + 1)] \cdot \int_{(2c-1)z/[6(c^2-c+1)]^{1/2}}^{(c+1)z/(1-c)[6(c^2-c+1)]^{1/2}} (2\pi)^{-1/2} \exp\,(-t^2/2)\, dt. \tag{1}$$

With the help of [3], [4], and [5], $g(z)$ can be tabulated. Values of $g(z)$ are given in Table I for several values of the parameter c.

From the general form (1), several densities can be derived as particular cases. For instance, the value $c = 0$ leads to the density of the range as given by McKay

Received March 28, 1955; revised July 7, 1957.

[1] Part of a Ph.D. thesis presented to the Institute of Statistics, University of North Carolina. This research was supported in part by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

TABLE I

Values of $g(z)$, for various values of the constant c, where
$z = x_{(0)} - cx_{(1)} - (1 - c)x_{(2)}$

z \ c	0	0 1	0 2	0 4	0 6	0.8	0 9	1 0
0.0	0.00000	0.00000	0.00000	0.00000	0 00000	0.00000	0.00000	0.84628
0.2	.10917	.12101	.13709	.17898	.26659	49282	.73839	.78334
0.4	.21095	.23318	.26050	.33877	.47562	.70969	.75554	.70763
0.6	.29932	.32877	.36410	45941	59859	.71048	67281	.62378
0.8	.36927	.40194	.43988	53834	.61049	.63299	.58344	.53652
1.0	.41774	.44958	.48459	55897	60386	54116	.49344	45022
1.2	.44376	.47102	.49861	.54388	.53555	.45020	.40687	.36835
1.4	.44833	.46822	.48548	.49838	45187	36497	.32709	.29429
1.6	.43408	.44502	.45086	.43473	.36725	28832	25636	22920
1.8	.40476	.40647	.40149	.36270	.28937	22196	.19588	.17410
2.0	.36474	.35800	.34410	.29160	22168	16649	.14590	.12896
2.2	.31842	.30485	.28468	22659	.16529	12170	.10593	09315
2.4	.26981	.25145	.23115	.17064	.12000	.08668	07497	.06560
2.6	.22221	.20121	.17665	.12479	08484	.06016	05171	.04504
2.8	.17809	.15639	.13290	08871	.05840	04068	.03477	.03016
3.0	.13903	.11819	.09715	06135	.03918	02679	02278	.01968
3.2	.10580	.08692	.06904	04130	02556	01720	.01455	01252
3.4	.07853	.06225	.04775	.02707	01625	01076	00905	00777
3.6	.05690	.04345	.03215	01727	.01006	00656	00548	.00469
3.8	.04026	.02975	.02109	.01073	00606	00389	.00327	00277
4.0	.02782	.01971	.01348	.00650	00356	00225	.00188	.00159

and Pearson [2]; while the value $c = 0\ 5$ leads to the density of $v = x_{(0)} - (x_{(0)} + x_{(1)} + x_{(2)})/3$ as given by McKay [1]. The complexity of the expression for $g(z)$ increases rapidly with the number of variables; consequently, we will limit our presentation to the above mentioned case.

Non-null distribution of linear contrasts of order statistics. Here again, and for the same reasons, only the case of three variables will be presented. In order to get the non-null distribution of $z = x_{(0)} - cx_{(1)} - (1 - c)x_{(2)}$ the joint density of $x_{(0)}$, $x_{(1)}$ and $x_{(2)}$ must be used as a starting point It is of the form

$$g(x_{(0)}, x_{(1)}, x_{(2)}) = \frac{1}{(2\pi)^{3/2}} \exp\left[\frac{-\mu'\mu}{2}\right] \exp\left[\frac{-X'X}{2}\right] \sum{}^{*} \exp(\mu_i' X),$$

where

$$\mu = \begin{vmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{vmatrix}, \qquad X = \begin{vmatrix} x_0 \\ x_1 \\ x_2 \end{vmatrix}, \qquad \mu_i = \begin{vmatrix} \mu_{i_0} \\ \mu_{i_1} \\ \mu_{i_2} \end{vmatrix}$$

and $\sum^*$ stands for the summation over all the permutations i_0, i_1, i_2 of the numbers 0, 1 and 2. Introducing the contrast z with the appropriate transformation and integregating out the extra variables, one gets, after a few simplifications,

the following expression for the non-null density of z:

$$f(z) = \frac{1}{\sqrt{2\pi}\sqrt{2(c^2 - c + 1)}} \exp\left[-\frac{1}{2}(\mu'\mu - M^2/3)\right]$$

$$(2) \qquad \cdot \sum{}^{*} \left\{ \exp\left[\frac{-(z^2 - 2\gamma_1 z)}{4(c^2 - c + 1)}\right] \exp\left[\frac{(\gamma_1 + 2\gamma_2)^2}{12(c^2 - c + 1)}\right] \cdot \int_{[(2c-1)z-(\gamma_1+2\gamma_2)]/[6(c^2-c+1)]^{1/2}}^{[(c+1)z-(1-c)(\gamma_1+2\gamma_2)]/[(1-c)(6(c^2-c+1))^{1/2}]} (2\pi)^{-1/2} \exp(-t^2/2)\, dt \right\},$$

where $\gamma_1 = \mu_{i_0} - c\mu_{i_1} - (1 - c)\mu_{i_2}$, $\gamma_2 = -(1 - c)\mu_{i_0} + \mu_{i_1} - c\mu_{i_2}$, and $M = \mu_0 + \mu_1 + \mu_2$. It is easy to see, looking at (2), how much more complicated an expression for $f(z)$ can become in the case of several variables.

Many particular cases of interest have been considered, using expression (2) as a starting point. Only two cases are reported here. The first one corresponds to the hypothesis $H_1: \mu_0 = \delta$, $\mu_1 = \mu_2 = 0$, $\delta > 0$. Denoting by $f(z \mid H_1)$ the density of z under the hypothesis H_1, one gets

$$f(z \mid H_1) = \frac{1}{\sqrt{\pi(c^2 - c + 1)}} \exp[(-\delta^2/3)(g_1 + g_2 + g_3)],$$

where g_1, g_2 and g_3 are functions of z and of the parameters δ and c given by

$$g_1(z; \delta, c) = \exp[-(3z^2 - 6\delta z - (2c - 1)^2\delta^2)/12(c^2 - c + 1)]I_1(z; \delta, c),$$

$$g_2(z; \delta, c) = \exp[-(3z^2 + 6c\delta z - (2 - c)^2\delta^2)/12(c^2 - c + 1)]I_2(z; \delta, c),$$

$$g_3(z; \delta, c) = \exp[-(3z^2 + 6(1 - c)\delta z - (1 + c)^2\delta^2)/12(c^2 - c + 1)]I_3(z; \delta, c).$$

The functions I_1, I_2, and I_3 are given by

$$I_1 = \int_{\frac{(2c-1)z-(2c-1)\delta}{[6(c^2-c+1)]^{1/2}}}^{\frac{[(c+1)z-(1-c)(2c-1)\delta]}{(1-c)[6(c^2-c+1)]^{1/2}}} \frac{\exp(-t^2/2)}{(2\pi)^{1/2}}\, dt, \qquad I_2 = \int_{\frac{(2c-1)z-(2-c)\delta}{[6(c^2-c+1)]^{1/2}}}^{\frac{(c+1)z-(1-c)(2-c)\delta}{(1-c)[6(c^2-c+1)]^{1/2}}} \frac{\exp(-t^2/2)}{(2\pi)^{1/2}}\, dt$$

$$I_3 = \int_{\frac{(2c-1)z+(1+c)\delta}{[6(c^2-c+1)]^{1/2}}}^{\frac{(c+1)z+(1+c)(1-c)\delta}{(1-c)[6(c^2-c+1)]^{1/2}}} \frac{\exp(-t^2/2)}{(2\pi)^{1/2}}\, dt.$$

Table II contains the values of $f(z \mid H_1)$, in the case $\delta = 1$, for several values of the parameter c.

The case of equal spacing of the true means, i.e., the one corresponding to the hypothesis $H_2: \mu_0 = 2\delta$, $\mu_1 = \delta$, $\mu_2 = 0$, yields a slightly more complicated expression for $f(z \mid H_2)$. Table III contains some values of $f(z \mid H_2)$, in the particular case $\delta = 1$, for a few values of the parameter c.

TABLE II

Values of the density of $z = x_{(0)} - cx_{(1)} - (1 - c)x_{(2)}$ under the hypothesis $H_1 : \mu_0 = \delta = 1, \mu_1 = \mu_2 = 0$

zc	0	0.1	0.2	0.4	0.6	0.8	0.9	1.0
0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.69550
0.2	.07843	.08707	09783	12988	19255	36249	58377	65223
0.4	15340	.16984	.19015	24916	35644	56737	63842	60265
0.6	22169	.24434	27187	34847	47043	60539	58628	54862
0.8	.28049	30717	33879	42129	52678	56566	52858	49195
1.0	.32764	35584	38797	46387	53322	50692	46915	43436
1.2	.36168	.38868	41801	47716	51027	44557	40985	37744
1.4	.38200	40550	42904	46589	45370	38510	35211	32259
1.6	.38882	40689	42265	43486	39556	32714	29733	27008
1.8	38314	39449	40156	39238	33636	27293	24660	22357
2.0	36658	37079	36936	34034	27999	22364	20067	18002
2.2	.34126	.33851	32948	28770	22839	17958	16014	14373
2.4	.30954	30284	28546	23706	18255	14116	12632	11184
2.6	27387	26024	24117	19065	14288	—	—	
2.8	23387	.21939	19834	14980	—		—	—
3.0	.19953	.18053	15882	—	—	—	—	—
3.2	16449	14500	—	—	—	—	—	—
3.4	13256	—	—	—		.		

TABLE III

Values of the density of $z = x_{(0)} - cx_{(1)} - (1 - c)x_{(2)}$ under the hypothesis $H_2 : \mu_0 = 2\delta, \mu_1 = \delta, \mu_2 = 0, \delta = 1$

zc	0	0.1	0.2	0.4	0.6	0.8	0.9	1.0
0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	54317
0.2	.04056	03960	05069	06759	10137	20168	38016	52172
0.4	08109	09037	10133	13497	20146	37844	51800	49788
0.6	12140	13481	15152	20107	29509	47802	50288	47151
0.8	.16106	.17860	20039	26360	37311	49628	47401	44252
1.0	.19934	.22058	24647	31926	42638	47505	44270	41092
1.2	.23518	25927	28808	36377	45112	.44210	40851	37691
1.4	.26731	.29305	32298	.39430	44918	40526	37182	.34006
1.6	.29437	32027	.34906	40863	42718	.36591	33331	30372
1.8	31476	33859	36482	40664	39249	.32495	.29381	.26591
2.0	.32837	.34951	36940	.38986	35119	.28344	.25435	.22887
2.2	.33363	.35004	.36285	.36159	30690	23920	.21601	19316
2.4	.33070	.34123	34610	32502	26260	19934	.17977	.15978
2.6	31996	32398	.32083	.28390	.21793	—	—	—
2.8	.30177	.29972	.28927	24141	—	—	—	—
3.0	.27902	.27019	25383	—	—	—	—	—
3.2	.25165	.23777	—	—	—	—	—	—
3.4	.22185	—	—	—	—	—	—	—

REFERENCES

[1] A. T. McKay, "The Distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe," *Biometrika*, Vol. 27 (1935), pp. 466–471.
[2] A. T. McKay and E. S. Pearson, "A note on the distribution of range in samples of size n," *Biometrika*, Vol. 25 (1933), pp. 415–420.
[3] National Bureau of Standards, "Tables of the exponential function e^x," Applied Math. Series, No. 14, 1951.
[4] National Bureau of Standards, "Tables of normal probability functions," Applied Math. Series, No 23, 1953.
[5] K. Pearson, *Tables for Statisticians and Biometricians*, Cambridge University Press, Part I, 1930.
[6] J. St-Pierre and A. Zinger, "The null distribution of the difference between the two largest sample values," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 849–851.
[7] S. S. Wilks, "Order statistics," *Bull. Amer. Math. Soc.*, Vol. 54 (1948), pp. 6–50.

ADMISSIBLE ONE-SIDED TESTS FOR THE MEAN OF A RECTANGULAR DISTRIBUTION[1]

By J. W. Pratt

Harvard University

1: Theorem. *Suppose we have a sample of $n > 1$ independent observations from a uniform distribution with unknown mean θ and known range R. Suppose we wish to test $H_0:\theta \leqq \theta_0$ against $H_1:\theta > \theta_0$. Then an essentially complete class of admissible tests is the class $\mathfrak{A}$ of all tests of the following type. Let u be the minimum observation, v the maximum. Let $g(u)$ be a nonincreasing function of u such that $g(u) = \theta_0 + \frac{1}{2}R$ for $u < \theta_0 - \frac{1}{2}R$. Accept H_0 if and only if $v < g(u)$.*

2. Discussion. The two-sided problem has been treated by Allan Birnbaum [1]. He showed that, for testing $H_0':\theta = \theta_0$ against $H_1':\theta \neq \theta_0$, an essentially complete class of admissible tests is the class of all tests of the following type. Let $v(u)$ be a nondecreasing function of u. Accept H_0 if and only if $v > v(u)$ and $\theta_0 - \frac{1}{2}R < u < v < \theta_0 + \frac{1}{2}R$.

Birnbaum [1] also noted that there is a uniformly most powerful size α test of $H_0':\theta = \theta_0$ against $H_1:\theta > \theta_0$, namely that accepting H_0' if $\theta_0 - \frac{1}{2}R < u < \theta_0 + (\frac{1}{2} - \alpha^{1/n})R$ and $v < \theta_0 + \frac{1}{2}R$. This corresponds in our notation to

$$g(u) = \begin{cases} \theta_0 + \frac{1}{2}R \text{ for } u < \theta_0 + (\frac{1}{2} - \alpha^{1/n})R, \\ \theta_0 - \frac{1}{2}R \text{ (say) otherwise.} \end{cases}$$

Received July 15, 1957; revised March 21, 1958.
[1] Research carried out at Stanford University and revisions at University of Chicago, and at Harvard University, under partial sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

In this rather simple situation, then, an essentially complete class of admissible tests of the simple hypothesis against one-sided alternatives consists of the uniformly most powerful test (just described) for each significance level, but the class of admissible tests of the composite hypothesis against one-sided alternatives is very general. The class of admissible tests of the simple hypothesis against two-sided alternatives is also very general, but quite different. It includes unions of admissible lower and upper one-sided rejection regions (if and) only if they are admissible for the simple hypothesis, and such unions form a portion "of measure zero" in the whole class.

In the following section we will prove the result stated in the first paragraph. The proof uses no general results of decision theory, such as the complete class theorem, but only direct methods of an essentially elementary constructive type. It obviously works in some slightly more general situations, which are given explicitly in [2].

3. Proof. Without loss of generality we may take $\theta_0 = 0, R = 2$. Since (u, v) is a sufficient statistic, an essentially complete class of tests is the class of all randomized tests based on (u, v). Suppose such a test is given, accepting H_1 with probability $\phi_0(u, v)$ when (u, v) is observed.

The triangle $T(\theta) = \{(u, v): \theta - 1 < u \leqq v < \theta + 1\}$ contains (u, v) with probability one if θ is the true mean. The probability of accepting H_0 using the test function ϕ is

$$E_\theta(\phi) = \iint_{T(\theta)} \phi(u, v) 2^{-n} n(n - 1)(v - u)^{n-2}\, du\, dv. \tag{1}$$

If $\theta \geqq 0$, then $u > -1$ with probability one If $\theta \leqq 0$, then $v < 1$ with probability one. Thus if (u, v) is not in $T(0)$, we know which hypothesis is correct. Accordingly, let

$$\phi_1(u, v) = \begin{cases} \phi_0(u, v) & \text{if} \quad (u, v) \in T(0), \\ 1 & \text{if} \quad u \leqq -1, \\ 0 & \text{if} \quad v \geqq 1. \end{cases} \tag{2}$$

Then ϕ_1 dominates ϕ_0, i.e. ϕ_1 is at least as good as ϕ_0 for any θ, i.e.

$$E_\theta(\phi_1) \gtreqless E_\theta(\phi_0) \quad \text{for} \quad \theta \lesseqgtr 0. \tag{3}$$

Define $f(v)$ for $-1 < v < 1$ by

$$\int_{-1}^{f(v)} (v - u)^{n-2}\, du = \int_{-1}^{v} \phi_1(u, v)(v - u)^{n-2}\, du, \quad -1 \leqq f(v) \leqq v. \tag{4}$$

Let

$$\phi_2(u, v) = \begin{cases} 1 \text{ if } u \leqq f(v),\ -1 < v < 1, \text{ or if } v \leqq -1, \\ 0 \text{ if } u > f(v),\ -1 < v < 1, \text{ or if } v \geqq 1. \end{cases} \tag{5}$$

Then, with respect to the density $2^{-n}n(n-1)(v-u)^{n-2}du$, $v-1 < u < v$, ϕ_2 has the same mass as ϕ_1 on each horizontal line in the (u, v)-plane, but concentrates it as far to the left as possible. Furthermore, $\phi_2 = \phi_1$ except on $T(0)$. Therefore

$$(6) \qquad E_\theta(\phi_2) \underset{\leqq}{\overset{=}{\leqq}} E_\theta(\phi_1) \quad \text{for} \quad \theta \underset{>}{\overset{\leqq}{=}} 0.$$

Therefore ϕ_2 dominates ϕ_1.

Define $g(u)$ for $-1 < u < 1$ by

$$(7) \qquad \int_u^{g(u)} (v-u)^{n-2}\, dv = \int_u^1 \phi_2(u, v)(v-u)^{n-2}\, dv, \quad u < g(u) \leqq 1,$$

if the right-hand side is positive. If the right-hand side vanishes, or if $u \geqq 1$, let $g(u) = -1$. If $u \leqq -1$, let $g(u) = 1$. Let

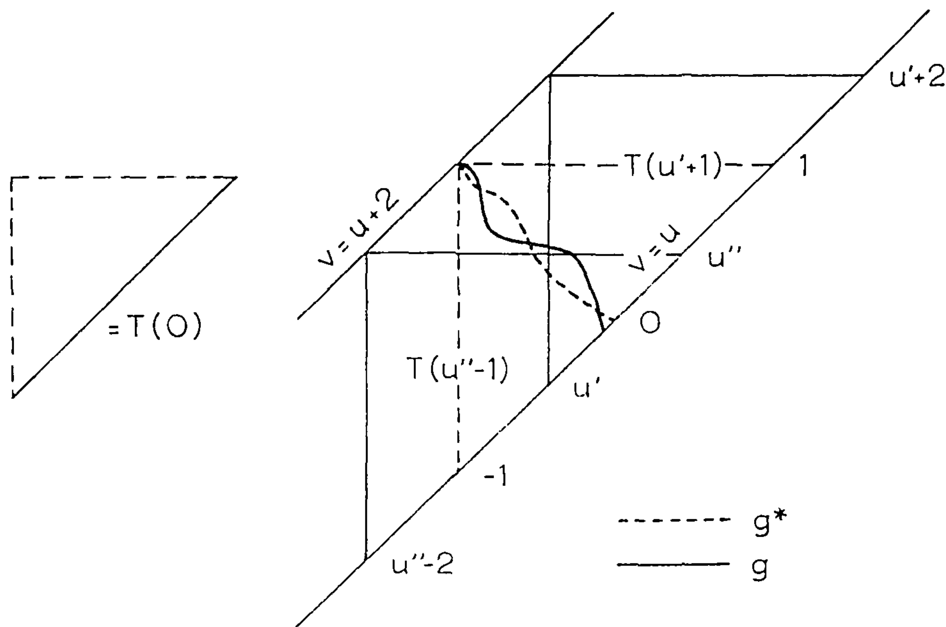
$$(8) \qquad \phi_3(u, v) = \begin{cases} 1 & \text{if} \quad v < g(u), \\ 0 & \text{if} \quad v \geqq g(u). \end{cases}$$

Then, with respect to the density $2^{-n}n(n-1)(v-u)^{n-2}\, dv$, $u < v < u+1$, ϕ_3 has the same mass as ϕ_2 on each vertical line in the (u, v)-plane, but concentrates it as low as possible. Furthermore, $\phi_3 = \phi_2$ except on $T(0)$. Therefore

$$(9) \qquad E_\theta(\phi_3) \underset{=}{\overset{\geqq}{=}} E_\theta(\phi_2) \quad \text{for} \quad \theta \underset{\geqq}{\overset{<}{=}} 0.$$

Therefore ϕ_3 dominates ϕ_2.

By (5), $\phi_2(u, v)$ is nonincreasing in u for each v. Therefore, by (7), for $-1 < u < g(u)$, $-1 < g(u) \leqq 1$, and $g(u)$ is nonincreasing in u. This is the essential part of the requirement that ϕ_3 be in $\mathfrak{A}$, and $g(u)$ was defined for other values of u so that ϕ_3 actually is in $\mathfrak{A}$.



We have thus shown that any test is dominated by a test in $\mathfrak{A}$, i.e. that $\mathfrak{A}$ is essentially complete. It remains to prove admissibility. Suppose ϕ and ϕ^* are given by g and g^*. Without changing the characteristics of the tests, we may redefine g and g^* so that they are left-continuous and so that $g(u) = -1$ where $g(u) \leqq u$, and $g^*(u) = -1$ where $g^*(u) \leqq u$. Suppose there is a u' such that $g(u') > g^*(u')$. Choose u'' such that $g(u') > u'' > g^*(u')$. (See the diagram.) Let "area" be measured with respect to the density $2^{-n}n(n-1)(v-u)^{n-2}du\,dv$. By left-continuity, $g^*(u) < u$ for all u in an interval whose right endpoint is u'. Therefore either the "area" below g in $T(u'+1)$ is less than that below g^*, or the "area" below g in $T(u''-1)$ is greater than that below g^*. But the "area" below g in $T(\theta)$ is just $E_\theta(\phi)$. Thus either $E_{u'+1}(\phi) < E_{u'+1}(\phi^*)$ or $E_{u''-1}(\phi) > E_{u''-1}(\phi^*)$. But $u'+1 > 0$ and $u''-1 < 0$, so this shows ϕ doesn't dominate ϕ^*. Hence if ϕ dominates ϕ^*, $g(u') \leqq g^*(u')$ for all u'. But in this case either ϕ and ϕ^* are essentially the same or $E_\theta(\phi) < E_\theta(\phi^*)$ for sufficiently small positive θ. Therefore ϕ cannot dominate ϕ^*. Since ϕ and ϕ^* were arbitrary tests of the essentially complete class $\mathfrak{A}$, it follows that all tests in A are admissible

This proof of admissibility is spelled out analytically in [2]. The proof of essential completeness given there uses a general property possessed by the rectangular distribution.

REFERENCES

[1] A. BIRNBAUM, "Admissible tests for the mean of a rectangular distribution," *Ann Math. Stat.*, Vol. 25 (1954), pp. 157-161.

[2] J. W. PRATT, "Some results in the decision theory of one-parameter multivariate Polya type distributions," Stanford Technical Report No 37 (Oct 21, 1955), pp 45-50

A METHOD FOR SELECTING THE SIZE OF THE INITIAL SAMPLE IN STEIN'S TWO SAMPLE PROCEDURE

BY JACK MOSHMAN

Corporation for Economic and Industrial Research, Arlington 2, Virginia

1. Summary and Introduction. The use of an upper percentage point of the distribution of total sample size in conjunction with the expectation of the latter is proposed as a guide to the selection of the size of the initial sample when using some version of Stein's [5] two-sample procedure. It is a rapidly calculable function of the underlying population variance based on existing tables of the χ^2 distribution. A rule-of-thumb is proposed to be used in making the actual selection of initial sample size. It is a simple matter to investigate the nature of the percentage point for different values of the variance over a limited range;

Received July 22, 1957; revised December 2, 1957.

a recommended conservative choice when the variance is not known is the selection of a large initial sample.

Dantzig [2] proved the nonexistence of nontrivial tests of Student's hypothesis whose power was independent of the variance, a result extended by Stein to the general linear hypothesis. In the same paper Stein proposed a two-sample procedure whose power was independent of variance. The same two-sample method could be used to obtain a confidence interval for the mean of a normal distribution with predetermined length and confidence coefficient.

Stein gave no specifications for the choice of the initial sample size, but Seelbinder [4] suggested that it be selected to minimize the expectation of the total sample. In a recent paper, Bechhofer, Dunnett and Sobel [1] used Stein's procedure for another application, noting that the variance of the total sample size increased as the size of the first sample decreased.

An efficient choice of the size of the initial sample will hold the expectation of the sample small, and will further reduce the probability of an extremely large total sample. This note will explore the matter in further detail and show that an upper percentage point of the distribution of total sample size, when used in conjunction with the expectation, is a rapidly calculable guide to an efficient choice of the size of the first sample.

2. Basic theory. As developed by Stein, the two-sample procedure involves a preliminary, arbitrary choice of a positive integer N_0 and a number $z > 0$. The value of z will depend, when constructing a confidence interval of length $2L$ for the mean, on the precision of the estimate, i.e., the length of the interval, and its reliability, the confidence coefficient. Specifically, if $t_{n,\gamma}$ is the upper $100\,\gamma$ percentage point of Student's distribution with n degrees of freedom, one would take $z = L^2/t^2_{N_0-1,1-(\alpha/2)}$ to obtain a confidence coefficient $\geqq 1 - \alpha$.

A sample of N_0 observations is taken and $s^2 = \sum(x_i - \bar{x})^2/(N_0 - 1)$ is computed as an estimate of the unknown variance σ^2 with $n = N_0 - 1$ degrees of freedom. The total sample size, N, is then

$$N = \max\left(\left[\frac{s^2}{z}\right] + 1, N_0\right), \tag{1}$$

where $[t]$ is the *largest integer less than* t.

Hence it follows that

$$\text{Prob}\,(N = N_0) = \text{Prob}\left(\frac{s^2}{z} \leqq N_0\right) = \text{Prob}\left(\frac{ns^2}{\sigma^2} = \chi^2(n) \leqq \frac{nN_0 z}{\sigma^2}\right), \tag{2}$$

where $\chi^2(n)$ is distributed as χ^2 with n degrees of freedom. Furthermore, for integral $m > N_0$,

$$\begin{aligned} \text{Prob}\,(N = m) &= \text{Prob}\left(m - 1 < \frac{s^2}{z} \leqq m\right) \\ &= \text{Prob}\left(\frac{n(m-1)z}{\sigma^2} < \chi^2(n) \leqq \frac{nmz}{\sigma^2}\right). \end{aligned} \tag{3}$$

a recommended conservative choice when the variance is not known is the selection of a large initial sample.

Dantzig [2] proved the nonexistence of nontrivial tests of Student's hypothesis whose power was independent of the variance, a result extended by Stein to the general linear hypothesis. In the same paper Stein proposed a two-sample procedure whose power was independent of variance. The same two-sample method could be used to obtain a confidence interval for the mean of a normal distribution with predetermined length and confidence coefficient.

Stein gave no specifications for the choice of the initial sample size, but Seelbinder [4] suggested that it be selected to minimize the expectation of the total sample. In a recent paper, Bechhofer, Dunnett and Sobel [1] used Stein's procedure for another application, noting that the variance of the total sample size increased as the size of the first sample decreased.

An efficient choice of the size of the initial sample will hold the expectation of the sample small, and will further reduce the probability of an extremely large total sample. This note will explore the matter in further detail and show that an upper percentage point of the distribution of total sample size, when used in conjunction with the expectation, is a rapidly calculable guide to an efficient choice of the size of the first sample.

2. Basic theory. As developed by Stein, the two-sample procedure involves a preliminary, arbitrary choice of a positive integer N_0 and a number $z > 0$. The value of z will depend, when constructing a confidence interval of length $2L$ for the mean, on the precision of the estimate, i.e., the length of the interval, and its reliability, the confidence coefficient. Specifically, if $t_{n,\gamma}$ is the upper $100\,\gamma$ percentage point of Student's distribution with n degrees of freedom, one would take $z = L^2/t^2_{N_0-1,1-(\alpha/2)}$ to obtain a confidence coefficient $\geqq 1 - \alpha$.

A sample of N_0 observations is taken and $s^2 = \sum(x_i - \bar{x})^2/(N_0 - 1)$ is computed as an estimate of the unknown variance σ^2 with $n = N_0 - 1$ degrees of freedom. The total sample size, N, is then

$$N = \max\left(\left[\frac{s^2}{z}\right] + 1, N_0\right), \tag{1}$$

where $[t]$ is the *largest integer less than* t.

Hence it follows that

$$\text{Prob}\ (N = N_0) = \text{Prob}\left(\frac{s^2}{z} \leqq N_0\right) = \text{Prob}\left(\frac{ns^2}{\sigma^2} = \chi^2(n) \leqq \frac{nN_0 z}{\sigma^2}\right), \tag{2}$$

where $\chi^2(n)$ is distributed as χ^2 with n degrees of freedom. Furthermore, for integral $m > N_0$,

$$\begin{aligned} \text{Prob}\ (N = m) &= \text{Prob}\left(m - 1 < \frac{s^2}{z} \leqq m\right) \\ &= \text{Prob}\left(\frac{n(m-1)z}{\sigma^2} < \chi^2(n) \leqq \frac{nmz}{\sigma^2}\right). \end{aligned} \tag{3}$$

Therefore, letting $\lambda = z/\sigma^2$, one may easily show

$$E(N) = N_0 \operatorname{Prob}(\chi^2(n) < n\lambda N_0) + \frac{1}{\lambda} \operatorname{Prob}(\chi^2(n+2) > n\lambda N_0) + \theta_1 \operatorname{Prob}(\chi^2(n) > n\lambda N_0) \tag{4}$$

and

$$\begin{aligned}\operatorname{Var}(N) = {} & N_0^2 \operatorname{Prob}(\chi^2(n) < n\lambda N_0) + \frac{(n+2)}{n\lambda^2} \operatorname{Prob}(\chi^2(n+4) > n\lambda N_0) \\ & + \frac{2\theta_2}{\lambda} \operatorname{Prob}(\chi^2(n+2) > n\lambda N_0) + \theta_3 \operatorname{Prob}(\chi^2(n) > n\lambda N_0) - (E(N))^2,\end{aligned} \tag{5}$$

where $0 \leqq \theta_i \leqq 1$, $i = 1, 2, 3$.

Whereas (4) defines $E(N)$ within a maximum error of unity, (5) is not as useful inasmuch as the factor $1/\lambda$ may be, and frequently is, large.

Furthermore, it is somewhat difficult to translate Var (N) into working percentage points of the distribution of N. A more useful procedure is to calculate a given percentage point N_p of the distribution. This may be accomplished directly from (2) and (3). Define N_p as the smallest integer $\geqq N_0$ such that

$$\operatorname{Prob}(N \leqq N_p) = \sum_{m=N_0}^{N_p} \operatorname{Prob}(N = m) \geqq p. \tag{6}$$

But this is equivalent, if one writes $p_n(\chi^2)$ as the probability density function of $\chi^2(n)$, to setting

$$\int_0^{nN_p\lambda} p_n(\chi^2)\, d\chi^2 \geqq p \tag{7}$$

and letting N_p be chosen to satisfy (7), but not less than N_0. Thus

$$N_p = \max\left\{N_0, \left[\frac{1}{\lambda}\left(100p\text{th } \textit{percentage point of } \frac{\chi^2(n)}{n}\right)\right] \text{[illegible]}\right. \tag{8}$$

which is tabulated in Hald [3] for example. Note that the [illegible] points of $\chi^2(n)/n$ decreases monotonically as n increases. [illegible] chosen very large, one can be reasonably confident that [illegible] be necessary, but this is not an efficient procedure.

A rough, but objective, rule-of-thumb may be [illegible] sideration: Let $E(N \mid N_0^*)$ be the expectation of N [illegible] the 100pth percentile of N if $N_0 = N_0^*$. Define

$$P(N_0^*) = \int_0^{n\lambda E(N \mid N_0^*)} \text{[illegible]} \tag{9}$$

as the proportion of time N will not exceed [illegible] N_0 which minimizes $E(N)$, i.e.,

$$E(N \mid N_0^*) \leqq \text{[illegible]} \tag{10}$$

for all N_0. Now one might investigate alternative values of N_0 by considering

$$\Psi(N_0) = (1 - p)(N_p(\mathbf{N}_0) - N_p(N_0)) - (1 - P(\mathbf{N}_0))(E(N \mid N_0) - E(N \mid \mathbf{N}_0)) \tag{11}$$

and selecting N_0 as the integer for which $\Psi(N_0)$ is a maximum. In effect, (11) weights the expected changes in $E(N)$ and N_p by the probability of exceeding those values. It would be expected that p would be chosen independently from nonstatistical considerations.

3. Example. If one takes $\lambda = .1$, where in the ordinary application considered by Stein $n = N_0 - 1$, then $E(N)$ is a minimum for $N_0 = \mathbf{N}_0 = 3$. Values of $E(N \mid N_0)$ are tabulated in Table 1. It may be seen that $E(N \mid N_0)$ is fairly constant over a considerable range. The same table also contains $N_{.95}(N_0)$ which decreases sharply where $E(N \mid N_0)$ is relatively constant.

It may readily be verified from (9) that $P(N_0) = P(3) \approx .64$. Rapidly one may evaluate $\Psi(N_0)$ from (11), taking $p = .95$, and find that $\Psi(6) = .2686$ is the maximum. Hence the rule suggested specifies $N_0 = 6$ as the proper choice.

4. Discussion. When the variance is unknown, two alternatives exist. It may be feasible to express the length of the confidence interval desired as a proportion of σ; no difficulty then ensues since λ is specified. If L is specified absolutely, in most practical cases a range for σ is known. One can then investigate the distribution of N for various values of σ in this range and make a subsequent choice of N_0.

The procedures suggested in this note are particularly applicable to those situations where repeated sampling is not contemplated and/or there exists a physical reason for wanting to avoid excessively large samples. The latter situation may obtain where larger individual samples may entail the purchase of additional test equipment or require the supplementing of a regular interviewing staff by extra employees.

TABLE 1

Dependence of $E(N)$ and $N_{.95}$ on N_0

$\lambda = .1$

N_0	$E(N \mid N_0)$	$N_{.95}(N_0)$	N_0	$E(N \mid N_0)$	$N_{.95}(N_0)$
2	10.45	38.41	10	11.84	18.80
3	10.29	29.96	12	12.92	17.89
4	10.45	26.05	14	14.35	17.20
5	10.51	23.72	16	16.15	16.66
6	10.63	22.14	18	18.02	18.00
7	10.80	20.99	20	20.02	20.00
8	11.18	20.10	22	22.01	22.00
9	11.43	19.38	24	24.00	24.00

5. Acknowledgment. The author wishes to express his indebtedness to Professor G. E. Albert for many helpful suggestions made in the pursuance of this research.

REFERENCES

[1] R. E. BECHHOFER, C W. DUNNETT AND M. SOBEL, "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance," *Biometrika*, Vol 41 (1954), p 170

[2] G. B. DANTZIG, "On the non-existence of tests of 'Student's' hypothesis having power functions independent of σ," *Ann Math Stat*, Vol 11 (1940), p. 186

[3] ANDERS HALD, *Statistical Theory with Engineering Applications*, John Wiley & Sons, 1952.

[4] B. M. SEELBINDER, "On Stein's two-stage sampling scheme," *Ann Math Stat.*, Vol 24 (1953), p. 640.

[5] CHARLES STEIN, "A two-sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat*, Vol 16 (1945), p 243.

ON A PROBLEM IN MEASURE-SPACES

BY V. S. VARADARAJAN

Indian Statistical Institute, Calcutta

Summary. Let $\mathfrak{F}$ be the family of all random variables on a probability space Ω taking values from a separable and complete metric space X. In this paper we prove that $\mathfrak{F}$ is in a certain sense a closed family. More precisely, if $\{\xi_n\}$ is a sequence of X-valued random variables such that their probability distributions converge weakly to a probability distribution P on X, then there exists an X-valued random variable on Ω with distribution P. An example is also given which shows that the assumption of completeness of X cannot in general be dropped.

1. Preliminary remarks. In what follows $(\Omega, \mathcal{S}, \mu)$ is a probability space and X a separable metric space. We denote by $\mathfrak{B}$ the class of Borel subsets of X defined as the minimal σ-field containing all open subsets of X.

A map φ of Ω into X is called a random variable if it is measurable i.e., $\varphi^{-1}(A) \varepsilon \mathcal{S}$ for each $A \varepsilon \mathfrak{B}$. If φ is a random variable we define as its distribution the measure μ_φ on $\mathfrak{B}$ given by

$$\mu_\varphi(A) = \mu\{\varphi^{-1}(A)\}$$

for all $A \varepsilon \mathfrak{B}$. A given probability measure P on $\mathfrak{B}$ is said to be induced from Ω if there exists a random variable φ such that $P = \mu_\varphi$.

Suppose we are given a sequence $\{P_n\}$ of probability measures on $\mathfrak{B}$. We say that $\{P_n\}$ converges weakly to a probability measure P on $\mathfrak{B}$ ($P_n \Rightarrow P$ in symbols) if

$$\lim_{n\to\infty} \int_X g\, dP_n = \int_X g\, dP$$

Received November 12, 1956; revised May 19, 1958.

for every bounded continuous function g on X. In terms of subsets of X this is equivalent to

$$\limsup_{n\to\infty} P_n(C) \leqq P(C)$$

for every closed set $C \subset X$ ([1]). When X is the real line with the usual topology, this convergence is equivalent to the usual convergence of distributions.

2. The main theorem. In this section we state and prove the main theorem. Before doing it we prove a lemma.

LEMMA. *Let X be a separable and complete metric space and $(\Omega, \mathcal{S}, \mu)$ a nonatomic probability space* ([2] p. 168). *Then any probability measure on $\mathcal{B}$ can be induced from Ω.*

PROOF. Since X is a separable metric space, it can be imbedded homeomorphically into a countable product of unit intervals by a celebrated theorem of Urysohn ([3] p. 125). Since it is also complete, the image of X will be a G_δ by a theorem of Larentieff ([3] p. 207). X can thus be regarded as a Borel subset of a countable product of unit intervals. This implies however that X can be regarded as a Borel subset of the unit interval since the unit interval and the countable product of such intervals can be connected by an one-one map which is measurable both ways. It is thus sufficient to show that any probability measure on the unit interval can be induced from Ω. This however is a well-known result.

We now prove the main theorem.

THEOREM. Let X be a separable and complete metric space and $(\Omega, \mathcal{S}, \mu)$ an arbitrary probability space. If $\{\xi_n\}$ is a sequence of X-valued random variables such that $\mu_{\xi_n} \Rightarrow P$ as $n \to \infty$ where P is a probability measure on $\mathcal{B}$, there exists an X-valued random variable ξ such that $P = \mu_\xi$.

PROOF. Any measure space can be decomposed into its atomic and nonatomic components and in view of the previous lemma we can assume that there is no nonatomic component in Ω. We can thus write $\Omega = A_1 \cup A_2 \cup \cdots$ where (i) $A_i \cap A_j = \phi$ for $i \neq j$, (ii) each A_i is an atom of $(\Omega, \mathcal{S}, \mu)$, and (iii) $\mu(A_i) = c_i > 0$ for each i. The distribution $P_n(\equiv \mu_{\xi_n})$ is then atomic and (since X is separable) has mass concentrated in a countable set of points, say $\{a_{n1}, a_{n2}, \cdots\}$. $P_n[a_{ni}] = c_i$ for $i = 1, 2, \cdots$.

We first assert that for each i, the set $D_i = \{a_{1i}, a_{2i}, \cdots\}$ has compact closure. If not, then for some i_0, D_{i_0} has a subset which has no limit point and which is infinite. We can assume without losing generality that this subset is D_{i_0} itself and that all the a_{ni_0} are distinct. If then $D \subset D_{i_0}$ is any subset, then D is closed and from $P_n \Rightarrow P$ it follows that $P(D) \geqq \limsup_{n\to\infty} P_n(D)$. If D is infinite then, $\limsup_{n\to\infty} P_n(D) \geqq c_{i_0}$. Thus for any infinite subset $D \subset D_{i_0}$, $P(D) \geqq c_{i_0} > 0$ which is a contradiction.

Thus each D_i has compact closure. We can then, by the diagonal procedure choose a sequence $\{n_k\}$ of integers and points $a_1, a_2, \cdots$ of X such that

$$\lim_{k\to\infty} a_{n_k,i} = a_i$$

$i = 1, 2, \cdots$. Let ξ be the random variable with values $a_1, a_2, \cdots$ on the s $A_1, A_2, \cdots$. We complete the proof by showing that $P = \mu_\xi$. It is enough show that $P_{n_k} \Rightarrow \mu_\xi$. In fact for any bounded continuous g on X,

$$\int_X g\, dP_{n_k} = \sum_i c_i\, g(a_{n_k,i}) \rightarrow \sum_i c_i\, g(a_i) = \int_X g\, d\mu_\xi,$$

passage to the limit being justified as $\sum_i c_i g(a_{n_k,i})$ converges uniformly in k. is completes the proof of the theorem.

REMARKS. (1) Suppose X is any separable metric space and X^* its completion. e above theorem will still be true not for X but for X^* and ξ will now be X^*-ued. If then X has the property that as a subset of X^* it is measurable with pect to the completion of *every* measure on X^*, ξ can be reduced to an X-valued dom variable and the main theorem is true for such X. This is the case for tance when X is itself a Borel set in X^*. It is interesting to note that there are arable metric spaces X which have the above mentioned property in relation X^* but which are not complete under any metrization, for example, the set rationals with the relative real line topology.

2) It is to be noted that when (Ω, S, μ) is purely atomic the theorem is true h any separable X.

3) Suppose now $A_1, A_2, \cdots$ is a sequence of sets in S such that $\mu(A_n) \rightarrow \alpha$. ting $\xi_n = \chi_{A_n}$, the characteristic function of A_n, we find that $\mu_{\xi_n} \Rightarrow P$ where s the measure with masses α and $1 - \alpha$ at the points 1 and 0. The above orem then ensures the existence of $A \in S$ such that $\mu(A) = \alpha$; in other words t the range of μ is a closed subset of $[0, 1]$.

. **An example.** We construct an example to show that the theorem proved in tion 2 requires some such condition on X. We take for X a subset of $[0, 1]$ h that (i) $\mu^*(X) = 1$, $\mu_*(X) = 0$ where μ is Lebesgue measure and (ii) X con-s all points of the form $m/2^n$. For (Ω, S, μ) we take the unit interval with esgue measure. The Borel sets of X are precisely the intersections with X Borel subsets of $[0, 1]$. Lebesgue outer measure on $\mathcal{B}$ is now actually a measure r it, denoted by λ.

uppose now P_n is the measure on $\mathcal{B}$ with equal masses $1/2^n$ at the points 2^n $(m = 1, 2, \cdots 2^n)$. It is easy to verify that $P_n \Rightarrow \lambda$. Further each P_n is ially induced from Ω. We will now show that λ cannot be induced from Ω. uppose λ is induced by the map ξ. ξ is obviously a Borel measurable function $[0, 1]$ and hence by Lusin's theorem ([2]) p.243) we can find for each $\epsilon > 0$ mpact $K_\epsilon \subset [0, 1]$ such that (i) $\mu(K_\epsilon) > 1 - \epsilon$ and (ii) ξ restricted to K_ϵ is tinuous. If $M_\epsilon = \xi[K_\epsilon]$, then $M_\epsilon \subset X$ and is a compact subset of the real . Since λ is induced by ξ, $\lambda(M_\epsilon) > 1 - \epsilon$. But M_ϵ is a Borel set of the real line this shows that $\mu(M_\epsilon) > 1 - \epsilon$, contradicting the assumption that $\mu_*(X) = 0$. hus λ cannot be induced from Ω. This completes the discussion of the example.

for every bounded continuous function g on X. In terms of subsets of X this equivalent to

$$\limsup_{n\to\infty} P_n(C) \leqq P(C)$$

for every closed set $C \subset X$ ([1]). When X is the real line with the usual topolog this convergence is equivalent to the usual convergence of distributions.

2. The main theorem. In this section we state and prove the main theoren Before doing it we prove a lemma.

LEMMA. *Let X be a separable and complete metric space and $(\Omega, \mathcal{S}, \mu)$ a nonatom probability space* ([2] p. 168). *Then any probability measure on $\mathcal{B}$ can be induce from Ω.*

PROOF. Since X is a separable metric space, it can be imbedded homeomorphi cally into a countable product of unit intervals by a celebrated theorem c Urysohn ([3] p. 125). Since it is also complete, the image of X will be a G_δ by theorem of Larentieff ([3] p. 207). X can thus be regarded as a Borel subset c a countable product of unit intervals. This implies however that X can be re garded as a Borel subset of the unit interval since the unit interval and the count able product of such intervals can be connected by an one-one map which i measurable both ways. It is thus sufficient to show that any probability measur on the unit interval can be induced from Ω. This however is a well-known result

We now prove the main theorem.

THEOREM. Let X be a separable and complete metric space and $(\Omega, \mathcal{S}, \mu)$ a arbitrary probability space. If $\{\xi_n\}$ is a sequence of X-valued random variable such that $\mu_{\xi_n} \Rightarrow P$ as $n \to \infty$ where P is a probability measure on $\mathcal{B}$, there exist an X-valued random variable ξ such that $P = \mu_\xi$.

PROOF. Any measure space can be decomposed into its atomic and nonatomi components and in view of the previous lemma we can assume that there is nc nonatomic component in Ω. We can thus write $\Omega = A_1 \cup A_2 \cup \cdots$ where (i) $A_i \cap A_j = \phi$ for $i \neq j$, (ii) each A_i is an atom of $(\Omega, \mathcal{S}, \mu)$, and (iii) $\mu(A_i) = c_i > 0$ for each i. The distribution $P_n (\equiv \mu_{\xi_n})$ is then atomic and (since X is separable) has mass concentrated in a countable set of points, say $\{a_{n1}, a_{n2}, \cdots\}$. $P_n[a_{ni}] = c_i$ for $i = 1, 2, \cdots$.

We first assert that for each i, the set $D_i = \{a_{1i}, a_{2i}, \cdots\}$ has compact closure. If not, then for some i_0, D_{i_0} has a subset which has no limit point and which is infinite. We can assume without losing generality that this subset is D_{i_0} itself and that all the a_{ni_0} are distinct. If then $D \subset D_{i_0}$ is any subset, then D is closed and from $P_n \Rightarrow P$ it follows that $P(D) \geqq \limsup_{n\to\infty} P_n(D)$. If D is infinite then, $\limsup_{n\to\infty} P_n(D) \geqq c_{i_0}$. Thus for any infinite subset $D \subset D_{i_0}$, $P(D) \geqq c_{i_0} > 0$ which is a contradiction.

Thus each D_i has compact closure. We can then, by the diagonal procedure choose a sequence $\{n_k\}$ of integers and points $a_1, a_2, \cdots$ of X such that

$$\lim_{k\to\infty} a_{n_k, i} = a_i$$

for $i = 1, 2, \cdots$. Let ξ be the random variable with values $a_1, a_2, \cdots$ on the sets $A_1, A_2, \cdots$. We complete the proof by showing that $P = \mu_\xi$. It is enough to show that $P_{n_k} \Rightarrow \mu_\xi$. In fact for any bounded continuous g on X,

$$\int_X g\, dP_{n_k} = \sum_i c_i\, g(a_{n_k, i}) \to \sum_i c_i\, g(a_i) = \int_X g\, d\mu_\xi,$$

the passage to the limit being justified as $\sum_i c_i g(a_{n_k, i})$ converges uniformly in k. This completes the proof of the theorem.

REMARKS. (1) Suppose X is any separable metric space and X^* its completion. The above theorem will still be true not for X but for X^* and ξ will now be X^*-valued. If then X has the property that as a subset of X^* it is measurable with respect to the completion of *every* measure on X^*, ξ can be reduced to an X-valued random variable and the main theorem is true for such X. This is the case for instance when X is itself a Borel set in X^*. It is interesting to note that there are separable metric spaces X which have the above mentioned property in relation to X^* but which are not complete under any metrization, for example, the set of rationals with the relative real line topology.

(2) It is to be noted that when (Ω, S, μ) is purely atomic the theorem is true with any separable X.

(3) Suppose now $A_1, A_2, \cdots$ is a sequence of sets in S such that $\mu(A_n) \to \alpha$. Setting $\xi_n = \chi_{A_n}$, the characteristic function of A_n, we find that $\mu_{\xi_n} \Rightarrow P$ where P is the measure with masses α and $1 - \alpha$ at the points 1 and 0. The above theorem then ensures the existence of $A \in S$ such that $\mu(A) = \alpha$; in other words that the range of μ is a closed subset of $[0, 1]$.

3. An example. We construct an example to show that the theorem proved in Section 2 requires some such condition on X. We take for X a subset of $[0, 1]$ such that (i) $\mu^*(X) = 1$, $\mu_*(X) = 0$ where μ is Lebesgue measure and (ii) X contains all points of the form $m/2^n$. For (Ω, S, μ) we take the unit interval with Lebesgue measure. The Borel sets of X are precisely the intersections with X of Borel subsets of $[0, 1]$. Lebesgue outer measure on $\mathcal{B}$ is now actually a measure over it, denoted by λ.

Suppose now P_n is the measure on $\mathcal{B}$ with equal masses $1/2^n$ at the points $m/2^n$ $(m = 1, 2, \cdots 2^n)$. It is easy to verify that $P_n \Rightarrow \lambda$. Further each P_n is trivially induced from Ω. We will now show that λ cannot be induced from Ω.

Suppose λ is induced by the map ξ. ξ is obviously a Borel measurable function on $[0, 1]$ and hence by Lusin's theorem ([2]) p.243) we can find for each $\epsilon > 0$ a compact $K_\epsilon \subset [0, 1]$ such that (i) $\mu(K_\epsilon) > 1 - \epsilon$ and (ii) ξ restricted to K_ϵ is continuous. If $M_\epsilon = \xi[K_\epsilon]$, then $M_\epsilon \subset X$ and is a compact subset of the real line. Since λ is induced by ξ, $\lambda(M_\epsilon) > 1 - \epsilon$. But M_ϵ is a Borel set of the real line and this shows that $\mu(M_\epsilon) > 1 - \epsilon$, contradicting the assumption that $\mu_*(X) = 0$. Thus λ cannot be induced from Ω. This completes the discussion of the example.

REFERENCES

[1] P. BILLINGSLEY, "The invariance principle for dependent random variables," *Trans. Amer. Math. Soc.* (1956).

[2] P. R. HALMOS, *Measure Theory*, D. Van Nostrand, New York, 1950.

[3] J. L. KELLEY, *General Topology*, D. Van Nostrand, New York, 1955.

CORRECTION TO "PROBABILITIES OF HYPOTHESES AND INFORMATION-STATISTICS IN SAMPLING FROM EXPONENTIAL-CLASS POPULATIONS"

BY MORTON KUPPERMAN

The George Washington University

In the paper cited in the title (*Ann. Math. Stat.*, Vol. 29 (1958), pp. 571–575): p. 572, line 5. For $\sum xp(x, \theta_m)$ read $\sum_x p(x, \theta_m)$.

CORRECTION TO "POWER FUNCTIONS OF THE GAMMA DISTRIBUTION"

G. D. BERNDT

Professor I. R. Savage has called to my attention, through the Editor, the fact that I have overlooked reference to previous work appearing in Eisenhart, Haystay, and Wallis, *Techniques of Statistical Analysis*, and bearing on results reported by me in the *Annals*, Vol. 29, No. 1, March 1958, pages 302–306.

On pages 274–275 of Eisenhart, Haystay, and Wallis, in Figures 8.1 and 8.2, there are given operating characteristic curves for the chi-squared distribution for eight selected degrees of freedom when the significance level is 0.01 and 0.05. Inasmuch as the chi-squared distribution is a gamma distribution with $\frac{1}{2}$ (degrees of freedom) = the parameter gamma in my paper and with 2 = the parameter beta in my paper, and since their rho is equivalent to my delta, there is a similarity in the reported results. This similarity has resulted in some overlap in the results of the two papers in that ten of my forty-eight power curves have an equivalent in the operating characteristic curves in the previous work.

I should like to acknowledge this previous work, and also that of Ferris, Grubbs, and Weaver, by having the following two references added to the two which already appear at the end of my paper:

[3] *Selected Techniques of Statistical Analysis*, Churchill Eisenhart, Millard W. Haystay, and W. Allen Wallis, editors, McGraw-Hill, New York, 1947, pp. 270–278.

[4] CHARLES D. FERRIS, FRANK E. GRUBBS, AND CHALMERS L. WEAVER, "Operating characteristics for some common statistical tests of significance," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 178–197.

ABSTRACTS OF PAPERS

(Abstracts of papers presented for the Cambridge, Massachusetts Meeting of the Institute, August 25–28, 1958.)

48. On the Relationship Algebra and the Association Algebra of the Partially Balanced Incomplete Block Design. JUNJIRO OGAWA, University of North Carolina.

A. T. James (1957) defined the so-called "relationship algebra of a design" and showed that the partition of the total sum of squares into partial sums of squares can be characterized by the structure of the relationship algebra. He constructed the relationship algebras and analyzed their algebraic structures for the randomized block, the latin-square and the balanced incomplete block design. The purpose of this note is to construct the relationship algebra for the partially balanced incomplete block design and analize its algebraic structure. The main result of this paper is that the second degree irreducible representations of the relationship algebra are completely determined by the irreducible (linear) representations of the association algebra defined by R. C. Bose. (Received June 2, 1958.)

49. Estimation of the Medians for Dependent Variables. OLIVE JEAN DUNN, University of California.

The problem considered in this paper is that of using non parametric methods to estimate the unknown medians of two dependent variables. In various types of research, it is convenient to consider a sample of n individuals and to take measurements at two different times or at two different levels of treatment. These $2n$ measurements are then a sample of size n from a bivariate distribution. For independent variables, two confidence intervals of the classic type using order statistics may be used as simultaneous confidence intervals for the two medians by simply multiplying the two probabilities to obtain the new confidence level. In this paper it is shown that for two dependent variables these same confidence intervals may be used as a set with bounded confidence level. Comparisons are made on the basis of average length between these intervals and other joint intervals for the means of a bivariate normal distribution. It is also shown that the result of this paper does not generalize, at any rate in the most obvious way, to three or more dependent variables. (Received June 16, 1958.)

50. On the Problem of Incomplete Data. (Preliminary report) JUNJIRO OGAWA and BERNARD S. PASTERNACK, University of North Carolina.

Consider a sample of size n, $x_1, x_2, \cdots, x_n$, drawn from $N(\mu, \sigma^2)$ (σ^2 known). Suppose only $n - k$ observations are available, $x_1, x_2, \cdots, x_{n-k}$ (say). Let

$$\bar{x} = 1/n \sum_{i=1}^{n} x_i, \ \bar{x}^* = 1/(n-k) \sum_{i=1}^{n-k} x_i, \ u = (\sqrt{n}\,\bar{x})/\sigma, \ u^* = (\sqrt{n-k}\,\bar{x}^*)/\sigma$$

and define $u(\alpha)$ by $P_{H_0}(|u| \leq u(\alpha)) = 1 - \alpha$ and $u^*(\alpha)$ by $P_{H_0}(|u^*| \leq u^*(\alpha)) = 1 - \alpha$, where $H_0{:}\mu = \mu_0$. We define $P_H\{|u| \leq u(\alpha) \mid |u^*| > u^*(\alpha)\}$, α being prefixed, as the *reversal function* of this test procedure. The reversal function has been tabulated for various values of k/n. When σ^2 is unknown, the test procedure depends upon t, or, for incomplete data, t^* [illegible] obtained, i.e., the minimum value of r such [illegible] l.u.b. and g.l.b.), in probability [illegible] have also been obtained for [illegible] classifications, the general linear model and Hotelling's T^2. Another approach to the problem of missing data involves the

introduction of a chance mechanism according to which observations are missed. Resea along these lines is now in progress and the authors hope to present some of these resu in the near future. (Received June 20, 1958.)

51. Aids for Fitting the Pearson Type III Curve by Maximum Likeliho (Preliminary report) J. ARTHUR GREENWOOD, Iowa State College a DAVID DURAND, M.I.T.

New tables and formulas of approximation are given for the function $\rho = y\phi(y)$, wh $\phi(y)$ is the inverse function to $y = \ln\rho - d/d\rho \ln\Gamma(\rho)$. With the aid of these tables, may obtain by direct interpolation the maximum likelihood estimate (joint) of the ponent in a Type III distribution with known lower limit. Application of the tables the Type III with unknown lower limit and to the Type V are briefly discussed. (Recei June 20, 1958.)

52. Admissible Estimates and Maximum Likelihood Estimates (Prelimin report) ALLAN BIRNBAUM, Columbia University.

A definition of admissibility of a point-estimate of a real-valued parameter θ is form lated on the basis of a slightly generalized form of the Neyman-Pearson theory of co dence regions, using *Ann. Math. Stat.*, Vol. 27 (1956), pp. 544–545, without introduct of loss functions. Necessary and sufficient conditions for existence of such estimates given under mild regularity conditions. By extending methods developed in *Ann. Ma Stat.*, Vol. 26 (1955), pp. 21–36, it is shown that each admissible estimate is obtaina as the (unique) solution of an equation $\partial/\partial\theta \log L(x, \theta) = G(\theta)$, where $G(\theta)$ is a kno function and $L(x, \theta)$ is the likelihood function. Setting $G(\theta) = 0$ gives the maximum lik hood estimate $\hat{\theta}$, which is thus shown to be admissible. In the case of non-existence admissible estimates, asymptotically admissible estimates are defined and shown un certain conditions to exist and to include $\hat{\theta}$. An estimate $\tilde{\theta}$ is called median-unbiased Prob $\{\tilde{\theta} \leqq \theta \mid \theta\} = \frac{1}{2}$ for all θ. $\hat{\theta}$ is shown under general conditions to be asymptotica median-unbiased, and to be a convenient approximation (often close for moderate sam sizes) to the median-unbiased admissible estimate (which is often difficult to comput Relations to sufficiency and to multi-parameter estimation problems are discussed. (R ceived June 24, 1958.)

53. Stochastic Models for the Electron Multiplier Tube (Preliminary repo EDWARD K. DALTON, WILLARD D. JAMES AND HOWARD G. TUCKE University of California.

Four stochastic models are proposed for the electron multiplier tube, two being branc ing processes involving Poisson distributions and two being branching processes involvi binomial distributions. In each model there are two unknown parameters. It is desir to determine the best model among these and to estimate the parameters for it. Althou the probability generating functions in each case are easy to derive, explicit formul for the probability distributions in each case could not be found. A method for testi these models is presented which is based on the following theorem. THEOREM. *Let X a random variable which takes on non-negative integer values, and let $X_1, \cdots, X_n, \cdots$ d note an infinite sequence of independent observations on X. Let $g(u \mid \alpha) = E(u^X)$ be the prob bility generating function of X which depends on a (vector) parameter α and is continuo in α. Let α_0 be the true value of α, and assume that there exists a sequence $\{\hat{\alpha}_n\}$ of rand variables which converges to α_0 with probability one. Then for any value of u f which $u^2 - u \neq 0$ and $g(u^2 \mid \alpha) < \infty$ there exists a subsequence $\{\hat{\alpha}_{N_n}\}$ of $\{\hat{\alpha}_n\}$ such that t*

limiting distribution of the ratio of $\sum\{u^{X_k} \mid 1 \leqq k \leqq n\} - ng(u \mid \hat{a}_{N_n})$ *to either the square root of* $n\{g(u^2 \mid \hat{a}_{N_n}) - g^2(u \mid \hat{a}_{N_n})\}$ *or to the square root of*

$$n(n^{-1}\sum\{u^{2X_k} \mid 1 \leqq k \leqq n\} - (n^{-1}\sum\{u^{X_k} \mid 1 \leqq k \leqq n\})^2)$$

is normal with mean zero and variance one. A resumé of numerical results is included for three different sets of data corresponding to three different energy inputs (Received June 26, 1958.)

54. On the Choice of Sample Size in the Kolmogorov-Smirnov Tests. JUDAH ROSENBLATT, Purdue University.

If F_n is the empirical distribution based on independent random variables $X_1, \cdots, X_n$, with common c.d.f. F, it is well known that a test of the hypothesis H_0 $F = F_0$ having asymptotic probability of type one error not exceeding α is to reject H_0 if and only if $n^{1/2} d_1(F_n, F_0) \equiv n^{1/2} \sup_x | F_n(x) - F_0(x) | > h_{1\alpha}$, where

$$\lim_{n\to\infty} P_F\{n^{1/2} d_1(F_n, F) > h_{1\alpha}\} = \alpha$$

if F is continuous. Massey has suggested that the sample size n needed to achieve

$$P_F\{\text{Reject } H_0\} \geqq 1 - \beta \text{ when } d_1(F_0, F) \geqq l$$

be chosen as follows: n is the smallest integer such that $2[n^{1/2} l - h_{1\alpha}] \geqq c_\beta$, where

$$\int_{-\infty}^{c_\beta} (1/2\pi)^{1/2} e^{-t^2/2}\, dt = 1 - \beta$$

This suggestion is motivated by the normal approximation to the binomial distribution A thorough investigation is made of this suggested procedure, and a completely justified, still rather simple technique is devised for choosing n such that

$$P_F\{\text{Reject } H_0\} \geqq 1 - \beta \text{ when } d_1(F_0, F) \geqq l.$$

The investigation is in two parts. First a region (near $p = \frac{1}{2}$) is determined where

$$\sum_{\nu=0}^{[n(p+l)-n^{1/2}h_{1\alpha}]} \binom{n}{\nu} p^\nu (1-p)^{n-\nu}$$

takes on its minimum value. This, together with the Uspensky version of the normal approximation to the binomial (with correction and error term) leads to the justified pro-

(Received July 2, 1958.)

55. The Use of Sample Quasi-Ranges in Estimating Population Standard Deviation. H. LEON HARTER, Wright Air Development Center.

The use of sample quasi-ranges in estimating the standard deviation of normal, rectangular, and exponential populations is discussed. For the normal population, the expected value, variance, and standard deviation of the rth quasi-range for samples of size n are tabulated for $r = 0(1)8$ and $n = (2r + 2)(1)100$. The efficiency of the unbiased estimate of population standard deviation based on one sample quasi-range is tabulated for the same values of r and n. Estimates based on a linear combination of two quasi-ranges are c [illegible] maxi-
mize [illegible] nge for
$n =$ [illegible] quasi-

ranges ($r < r' \leqq 8$) for $n = 4(1)100$ are tabulated, along with their efficiencies. An example illustrates the use of these estimates. For rectangular and exponential populations, the most efficient unbiased estimates based on one quasi-range are tabulated, together with their efficiencies, also the bias when estimates which assume normality are used. (Received July 2, 1958.)

56. On a Limiting Distribution Due to Renyi. D. G. Chapman, University of Washington.

Let X be a real valued random variable with distribution function (d.f.) $F(x)$. Let $F_n(x)$ denote the emperical d.f. based on n independent observations $x_1, x_2, \cdots, x_n$ of X. Renyi ("On the theory of order statistics," *Acta Math.*, Acad. Sci. Hungary, Vol. 4 (1953), pp. 191–231) has given the limiting distribution of $n^{1/2} R_n(a) = n^{1/2} \sup_{F(x) \geqq a} [F_n(x) - F(x)]/F(x)$ as n tends to infinity, a being an arbitrary positive constant. It is therefore of interest to determine the limiting distribution of $R_n(0)$, i.e., without the arbitrary restriction $F(x) \geqq a$. The result is obtained that $P_r[R_n(0) \leqq \epsilon] = \epsilon/1 + \epsilon$ for all n, so that the limiting distribution of $R_n(0)$ has the same form. Also studied in this paper are the limiting distributions of some slight generalizations of $R_n(a)$. The method used is that due to Doob which is simpler than Renyi's and may also be used to determine the asymptotic power of the Smirnov test of goodness-of-fit for certain alternatives. (Received July 2, 1958.)

57. Power and Control of Size of Some Optimal Welch-type Statistics. Roger S. McCullough and John Gurland.

A Welch-type statistic (Welch, *Biometrika*, 1938) is considered for testing equality of means in two normal populations with unknown variances which may be unequal. For various combinations of small sample sizes a nearly perfect one-sided control of size is possible, that is, optimal statistics are available which keep the size extremely close to a preassigned level if one population has a larger variance than the other. For two-sided control of size, that is with no restriction on the direction of inequality of variances, optimal statistics are available which keep the size below a pre-assigned level but arbitrarily close to the level over an infinite range of variance values. A table giving the optimal statistics for various combinations of small sample sizes has been prepared with the aid of an electronic computer. Tables of the power are also included. (Received July 2, 1958.)

58. A Note on Estimating Translation and Scalar Parameters. Joseph A. Dubay, University of Oregon.

Let $X = (X_1, \cdots, X_n)$ be a random variable whose distribution depends on an unknown real valued parameter θ. Let $\delta(X)$ be an estimator of θ, Γ be the class of all maximal translation invariant functions of X and assume the loss in estimating θ by $\delta(X)$ is $k(\delta(X)-\theta)^2$. A necessary and sufficient condition that among all estimators of the form $\delta(X) + u\gamma(X)$, $\gamma \varepsilon \Gamma$, u constant, $\delta(X)$ uniquely minimize the risk is given and an explicit construction of the minimum risk estimator is derived therefrom. In the particular case where $\delta(X)$ has the translation property, the class of estimators of the form $\delta(X) + u\gamma(X)$ is the class of all estimators having the translation property. Thus, a construction of the minimum risk estimator having the translation property is exhibited of which the constructions given by Pitman (1939) and Blackwell and Girshick (1954) in the case where θ is a translation parameter are special cases. An example is given in which θ is not a translation parameter in the usual sense but estimators having the translation property are naturally admitted. Under an appropriate transformation the results are applicable to the estimation of scalar parameters. (Received July 2, 1958.)

of such processes); (2) models based on the conditional probability of failure function; (3) extreme value models. Implications of and interrelations among the various models are discussed. Many examples are given. As examples of models (1) and (2) one may cite the recent paper by Z. W. Birnbaum and S. C. Saunders (*J. Amer. Stat. Assn.*, Vol. 53 (1958), pp. 151–160) in which they give a statistical model for the life length of structures under dynamic loading (i.e., fatigue) and a recent report by George H. Weiss in which it is shown that some kinds of mechanical failure, such as creep failure of oriented polymenic filaments under tensile stresses, can be viewed as "pure death" processes. An example of where model (3) may be relevant is in phenomena involving corrosion. (Received July 2, 1958.)

64. Truncation and Tests of Hypotheses. OM P. AGGARWAL AND IRWIN GUTTMAN, Purdue University and Princeton University.

Consider a normal distribution with variance σ^2 and a sample from the distribution obtained from this normal distribution by truncating it at the same distance a on both sides of the mean. The distribution of the sample mean for sample sizes up to 4 is obtained explicitly and the results of applying the usual tests of hypotheses for one-sided testing of the mean of a normal distribution are examined when a and σ^2 are known. Some tables are given and it is found that the loss in power decreases very rapidly with the distance of the alternative value of the mean from the one tested and also with the distance of the truncation from the mean. (Received July 2, 1958.)

65. Mathematical Outline of Polyvariable Analysis (Including Random Balance). F. E. SATTERTHWAITE, Statistical Engineering Institute.

A polyvariable technique for statistical analysis is defined as any estimation procedure applied to the linear model, $Y = BZ + E = BZ + EI = AX$, $A = (B, E)$, $X = (Z, I)$, which gives estimates for all (or of some) of the A unknowns with associated confidence limits that are *valid* and *finite* without restrictions on the number of A unknowns in the model. Specifically the number of unknowns may exceed, and often will greatly exceed, the number of data sets. The theoretical minimum number of data sets is 2. The necessary minimum for a specific application to give useful precision depends primarily on the signal-noise ratio for the available data. In many types of applications satisfactory precisions are obtained with 5 to 30 data sets for models containing a large number of unknowns. This paper is a mathematical outline of method and justification (including, in most cases, formal proofs) for the more important classes of polyvariable methods: (1) Polygression, (2) Bigression, (3) Quadratic, (4) Homovariance, (5) Hetervariance, (6) Random Balance, (7) Split Data. (Received July 3, 1958.)

66. Statistical Theory of Some Quantal Response Models. ALLAN BIRNBAUM, Columbia University. (By title)

Let $V = (S_1, \cdots S_k)$, where S_g's are independent Bernoulli observations: Prob $\{S_g = 1\} = P_g(y)$, a known strictly-increasing function of the unknown real-valued parameter y, Prob $\{S_g = 0\} = Q_g(y) = 1 - P_g(y)$, for $g = 1, \cdots k$. If $P_g(y)$ depends on known parameters a_g, b_g, $\cdots$, whose values the experimenter may determine, these are called "design parameters." Fisher's (*Phil. Trans. Roy. Soc. London*, A, Vol. 222(1922), pp. 363–366) method in treating estimation and design problems in the dilution series model ($P_g(y) = 1 - \exp(-a_g y)$, $g = 1, \cdots k$) is formulated more explicitly, particularly the use of the practical equivalence of designs having similar information curves $I(y) = \Sigma I_g(y)$, where $I_g(y) = (\partial/\partial y P_g(y))^2/P_g(y)Q_g(y)$. The "information area" $\int I(y)\, dy$ is introduced and

of such processes); (2) models based on the conditional probability of failure function; (3) extreme value models. Implications of and interrelations among the various models are discussed. Many examples are given. As examples of models (1) and (2) one may cite the recent paper by Z. W. Birnbaum and S. C. Saunders (*J. Amer. Stat. Assn.*, Vol. 53 (1958), pp. 151–160) in which they give a statistical model for the life length of structures under dynamic loading (i.e., fatigue) and a recent report by George H. Weiss in which it is shown that some kinds of mechanical failure, such as creep failure of oriented polymenic filaments under tensile stresses, can be viewed as "pure death" processes. An example of where model (3) may be relevant is in phenomena involving corrosion. (Received July 2, 1958.)

64. Truncation and Tests of Hypotheses. Om P. Aggarwal and Irwin Guttman, Purdue University and Princeton University.

Consider a normal distribution with variance σ^2 and a sample from the distribution obtained from this normal distribution by truncating it at the same distance a on both sides of the mean. The distribution of the sample mean for sample sizes up to 4 is obtained explicitly and the results of applying the usual tests of hypotheses for one-sided testing of the mean of a normal distribution are examined when a and σ^2 are known. Some tables are given and it is found that the loss in power decreases very rapidly with the distance of the alternative value of the mean from the one tested and also with the distance of the truncation from the mean. (Received July 2, 1958.)

65. Mathematical Outline of Polyvariable Analysis (Including Random Balance). F. E. Satterthwaite, Statistical Engineering Institute.

A polyvariable technique for statistical analysis is defined as any estimation procedure applied to the linear model, $Y = BZ + E = BZ + EI = AX$, $A = (B, E)$, $X = (Z, I)$, which gives estimates for all (or of some) of the A unknowns with associated confidence limits that are *valid* and *finite* without restrictions on the number of A unknowns in the model. Specifically the number of unknowns may exceed, and often will greatly exceed, the number of data sets. The theoretical minimum number of data sets is 2. The necessary minimum for a specific application to give useful precision depends primarily on the signal-noise ratio for the available data. In many types of applications satisfactory precisions are obtained with 5 to 30 data sets for models containing a large number of unknowns. This paper is a mathematical outline of method and justification (including, in most cases, formal proofs) for the more important classes of polyvariable methods: (1) Polygression, (2) Bigression, (3) Quadratic, (4) Homovariance, (5) Hetervariance, (6) Random Balance, (7) Split Data. (Received July 3, 1958.)

66. Statistical Theory of Some Quantal Response Models. Allan Birnbaum, Columbia University. (By title)

Let $V = (S_1, \cdots S_k)$, where S_g's are independent Bernoulli observations: Prob $\{S_g = 1\} = P_g(y)$, a known strictly-increasing function of the unknown real-valued parameter y, Prob $\{S_g = 0\} = Q_g(y) = 1 - P_g(y)$, for $g = 1, \cdots k$. If $P_g(y)$ depends on known parameters a_g, b_g, $\cdots$, whose values the experimenter may determine, these are called "design parameters." Fisher's (*Phil. Trans. Roy. Soc. London*, A, Vol. 222(1922), pp. 363–366) method in treating estimation and design problems in the dilution series model ($P_g(y) = 1 - \exp(-a_g y)$, $g = 1, \cdots k$) is formulated more explicitly, particularly the use of the practical equivalence of designs having similar information curves $I(y) = \sum I_g(y)$, where $I_g(y) = (\partial/\partial y P_g(y))^2/P_g(y)Q_g(y)$. The "information area" $\int I(y)\, dy$ is introduced and

9. Industrial Experience with Fractional Replicates. Cuthbert Daniel, (Invited paper)

Typical conditions of industrial experimentation (including numbers of factors si-
ıultaneously studied, number and magnitude of effects sought, and restrictions on time
nd costs) are reviewed. For meeting these, the sequential use of nested sub-fractions of
ractional replicate designs in the 2^{p-q} series is described. Since generally choice of a most
ıformative initial sub-fraction is incompatible with choice of a most informative com-
lete fractional replicate, the relative merits of each type, and of intermediate types, are
iscussed. A number of sequential designs are given. Methods are recommended and il-
ıstrated for inspection and criticism of data from 2^{p-q} experiments by using the graph
on appropriate probability paper) of the empirical c.d.f. of absolute values of contrasts,
o detect one or two mavericks, inadvertent plot-splitting, antilognormal data, and the
resence of several real effects. The distribution of this c.d.f. is studied under several hypoth-
ses, and the use is described of the operating characteristics of a related statistic given
y A. Birnbaum. Partial duplication is recommended when an unbiased estimate of error
ariance is required at an early stage. (Received July 7, 1958.)

0. On the Analysis of Factorial Experiments without Replication. Allan Birnbaum, Columbia University. (By title)

Inferences from factorial experiments without replication are usually based on a formal
ssumption that certain interactions are zero. In an altogether exploratory research situ-
tion, any statistical model giving a formal basis for informative inferences will typically
e too schematic and restrictive of unknown conditions to be claimed "valid," or a basis
or inferences which are "valid" except in the hypothetical formal sense; such a model is,
erhaps along with other models, a basis for "plausible inferences," i.e., inferences drawn
ı a formally-valid manner, based on a model which is more or less plausible. Under some
onditions (which are reviewed), the following schematic model is usefully plausible: The
ι contrasts a_i are independent, normal, homoscedastic; at most (any) r of their means
re non-zero. For $r = 1$, to decide which, if any, mean is non-zero, the statistic $\max_i a_i^2/\sum_i a_i^2$
; optimal. An alternative graphical procedure developed by C. Daniel, which has important
dvantages, is related to the ratio of $\max_i |a_i|$ to another ordered $|a_i|$. Critical values, power
ıd related properties, comparisons with more conventional statistics, and discussion of
ıses $r > 1$, are given. (Received July 7, 1958.)

1. Linear Regression in the Multivariate Normal Case. Charles Stein, University of California, Berkeley.

The problem of estimating the regression vector of one random variable on p others
hen all have a joint normal distribution is considered. There are $n \geqslant p + 2$ observations
ı the whole vector, the mean is assumed 0 for simplicity, and the loss is taken to be the
ıean squared error of prediction when the estimated regression vector is used to make a
rediction on the basis of a new random observation on the predictors, divided by the re-
dual variance. The usual (maximum likelihood) estimate of the regression vector is
ıinimax. It is admissible for $p = 1$, $n \geqslant 4$ and for $p = 2$ and n sufficiently large. For $p \geqslant 3$
; is intuitively clear (by analogy with the problem of estimating the mean of a multi-
ariate normal distribution) that the usual estimate is not admissible. One possible method
f improvement is to multiply the usual estimate by a constant depending on the popu-
ıtion multiple correlation coefficient, which can be estimated from the sample coefficient.
his will be more helpful if a guessed regression or a regression on a small selected set of
redictors is first subtracted out. Other possible improvements are suggested when the
ovariance matrix of the predictors is known. It should also be possible to make further

improvements when the covariance matrix of the predictors is not known but guessed or estimated on the basis of an additional sample. (Received September 4, 1958)

72. Some Population Estimation Models and Related Limit Distributions.

RONALD PYKE AND N. DONALD YLVISAKER, Stanford University.

The following two stage tag-and-sample model is studied During stage I, $J + 1$ samples of sizes $m_0, m_1, \cdots, m_J$ are taken from a population of size S_1 In each sample, all untagged members are tagged and the sample replaced. During a later stage II, $K + 1$ samples of sizes $(n_0, n_1, \cdots, n_K)$ are taken in each of which all tagged members are tagged with a different tag than that used in stage I. The time interval between stages is assumed to be large relative to the time required to perform the tagging and sampling Constant deterministic birth and death rates μ, and ρ are assumed during the intermediary time period Maximum Likelihood estimates of S_1, μ and ρ are obtained under both Poisson and Binomial assumptions on the distribution of the recovery random variables (r.v). Some gen-
able reparametrization
v 's and the Maximum
rther generalization in
which the sample sizes are assumed to be r v.'s is considered These results are then applied to data obtained from actual field experiments (Received July 7, 1958)

73. Applications of a certain Representation of the Wishart Matrix.

ROBERT A. WIJSMAN, University of Illinois

Let X be a $p \times n$ matrix $(p \leqq n)$ whose columns are independent and distributed like $N(0, \Sigma)$. It is known (e.g., J. G. Mauldon, *J Roy Stat. Soc*, Ser B, Vol. 17 (1955) pp 79–85) that the Wishart matrix XX' can be written as $CTT'C'$, where $CC' = \Sigma$, T is lower

T_{22} (this representation was also obtained by G. Elfving, *Skand. Aktuarietids*, Vol 30 (1947), pp. 56–74). This can be described as a non-central t_{n-1} variable, with random noncentrality parameter $T_{11}\,\rho/(1 - \rho^2)^{1/2}$. If the population multiple correlation between one variate and the correlation R can be represented by $R^2/(1$ $)/T^2_{pp}$. This is a non-central $F_{p-1,n-p+1}$ variab $R^2/(1 - R^2)$ The sphericity criterion Z (T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1958, section 10.7) in a bivariate population, when the hypothesis is true, can be represented by $Z/(1 - Z) = 2T_{11}T_{22}/((T_{11} - T_{22})^2 + T^2_{21})$, which is an $F_{2n-2,2}$ variable.
(Received July 7, 1958.)

74. Order Statistics and Estimation.

M. M. RAO, University of Minnesota. (Introduced by Milton Sobel) (by title)

Let (1) $f(x) = e^{-x}$, if $x > 0$, and zero otherwise, and X_i be the ith order statistic from a sample of N independent observations from the population defined by (1). The following results are proved. (I): Let $1 \leqq r_1 < r_2 < \cdots < r_p \leqq N$ be a set of fixed integers and $X_{r_1}, X_{r_2}, \cdots$ be a p set (subset) of the order statistics $X_1 < X_2 < \cdots < X_N$. Then
he parameter set, which
nsional of

$= \log \varphi_{r_1 \cdot r_2 \cdot \cdots \cdot r_p}(\xi_1, \xi_2, \cdots \xi_p) = \log N!/(N - r_p)! - \sum_{j=0}^{p-1} \sum_{m=r_j}^{r_{j+1}-1} \log (N - i\eta_{j+1} - m)$ where $\eta_j = \sum_{t=j}^{p} \xi_t$, $r_0 = 0$, and φ is the ch.f. (II): Let X_{r_i} and r_i be defined as in (I). Then $\{X_{r_i}, 1 \leqq r_i \leqq N\}$ forms a Markov Process in the strict sense as well as in the wide sense (in either case, the Process is non-stationary). (III): Some problems of interest (in physiological data) are the following: (i) The r_i are random but N is fixed. Suppose Prob $\{r_k = i_k, k = 1, 2, \cdots, p \mid r_i < r_{i+1}, i = 1, 2, \cdots p - 1\} = p_{i_1 \cdot i_2 \cdot \cdots \cdot i_p}$, and $p_{i_1 \cdot \cdots \cdot i_p} \geqq 0$, $\sum_{i_1 \cdot \cdots \cdot i_p} p_{i_1 \cdot \cdots \cdot i_p} = 1$, where $(i, < i_2 < \cdots < i_p) = 1, 2, \cdots N$, $i_0 = 0$, and the p's depend on a set of constants, $(\lambda_1, \lambda_2, \cdots, \lambda_k)$. Then the X_{r_i} defined similar to those in (I) still form a Stochastic Process whose finite-dimensional d.f.'s are determined by the ch.f. $\varphi(\xi_1, \cdots, \xi_p; \lambda_1, \cdots \lambda_k) = \sum_{i_1 \cdot \cdots \cdot i_p} p_{i_1 \cdot i_2 \cdot \cdots \cdot i_p} [\prod_{m=1}^{i_p} (N - m + 1)/\prod_{j=1}^{p} \prod_{m=i_{j-1}}^{i_{j+1}} (N - i\eta_j + m)]$. (ii) The case when N is a random variable. Specifying the d.f.'s in some cases of interest for the r_i, the limit d.f.'s of some linear combinations of X_i are considered. The estimation of the constants (A, θ) and the distribution of the $(\hat{A}, \hat{\theta})$ are treated using the above results when in (1) x is replaced by $(x - A)/\theta$, $\theta > 0$. (Received July 7, 1958.)

75. A Note on Order Statistics and Stochastic Independence.

GERALD S. ROGERS, University of Arizona.

The following theorem is proved. Let x be a continuous or discrete type real random variable. Let $x_1 \leqq \cdots \leqq x_n$ be the order statistics based on a random sample of size n from this x distribution. Let $z = z(x_1, \cdots, x_j)$ be a statistic based on the first $j < n$ items only. If z is stochastically independent of x_j, then z is stochastically independent of all x_k, $j < k \leqq n$; if z is stochastically independent of some x_k, $j < k \leqq n$, then z is stochastically independent of x_j and hence of all x_k, $j \leqq k \leqq n$. The first result is direct, since in terms of the conditional probability density functions, $g(z \mid x_j) = g(z \mid x_j, \cdots, x_n)$. For the second part, in $g(x_1, \cdots, x_{k-1} \mid x_k)$, let x_k be considered as a "parameter." Then, $(x_{k-1} \mid x_k)$ is a "complete single sufficient statistic" for x_k; also, the distribution of $(z \mid x_k)$ is free of the "parameter x_k." By a well known theorem, (Basu, *Sankhya*, Vol. 15 (1955), pp. 377–380), $(z \mid x_k)$ and $(x_{k-1} \mid x_k)$ are stochastically independent. It follows that z and x_{k-1} are stochastically independent; similarly, with an induction, z and x_k, $j \leqq k \leqq n$, are stochastically independent. (Received July 7, 1958.)

76. A model for Failure Data and its Applications. (Preliminary report)

ANDRE G. LAURENT, Wayne State University.

When a "ageing process" takes place, the response pattern of a "system" to a stimulus X does not follow an exponential distribution. The model $S(t) = \exp[1 + t - \exp(t)]$, where $S(t)$ is the "survival function," i.e., the integral of the "X-to-failure" distribution and $t = (X - X_0)/\tau$, has been proposed to meet this situation and tables provided for its use (*Oper. Res.*, February, 1957, p. 150; *Oper. Res. 13th National Meeting*, p. 35.) The present paper describes the more important features of the model above and gives the formulas for the expected values and the covariance matrix of the order statistics of a sample of size n. Tables of the expected values and the variances for $n = 1$ to 15, of the covariances for $n = 2$ to 5 are provided. The minimum variance linear unbiased estimates of the parameters of the distribution based on order statistics are studied for small samples and compared to other estimates from the viewpoint of efficiency. Related models are considered. (Received July 7, 1958.)

77. A Convolutive Class of Monotone Likelihood Ratio Families.

S. G. GHURYE AND DAVID L. WALLACE, University of Chicago.

A one-dimensional family $f(x, \theta)$ of densities on the real line or of probabilities on the integers, with the real parameter θ, is called a monotone likelihood ratio family if the ratio

$f(x, \theta')/f(x, \theta)$ is nondecreasing in x for $\theta \leq \theta'$. If several monotone likelihood ratio families each have all probability on two points which are the same for all families and all parameter values, then their convolution is a monotone likelihood ratio family. The extent to which similar results hold for distributions on three and more points and, with appropriate extensions of definitions, for multidimensional distributions on the vertices of the simplex and the cube is determined. A sufficient condition that the convolution of monotone likelihood ratio families be a monotone likelihood ratio family is that for each family, the ratio $f(x + h, \theta)/f(x, \theta)$ be non-increasing in x for all $h > 0$ (Received July 7, 1958.)

78. On the Exact Joint Distribution of the First Two Serial Correlation Coefficients.

V. K. MURTHY, University of North Carolina.

Any test of the hypothesis that up to a particular lag the true serial correlation coefficients are zero against some suitable alternative seems to necessitate knowledge of the joint distribution of serial correlation coefficients. As far as the author is aware even in the case of the first two serial correlation coefficients, the joint distribution has not so far been obtained in a simple closed form In this note the joint distribution of r_1 and r_2 has been obtained for samples of independent normal variates assuming the sample size to be of the form $4n + 1$ where n is a positive integer and adopting the circular definition suggested by Hotelling. This result has been obtained using a result of R L Anderson on the characteristic roots of the serial covariance, and inversion formulae for the distribution of ratios of quadratic forms given by Gurland Some properties of the joint distribution are obtained. The case of more than two serial correlation coefficients will be dealt with in a subsequent paper. (Research under ONR contract Nonr 855(06)) (Received July 7, 1958; revised July 28, 1958.)

79. Confidence Bounds Associated with a Test for Symmetry.

R. GNANADESIKAN, The Procter and Gamble Company.

In a p-variate nonsingular normal distribution $N[\mu, \Sigma]$, one may be interested in testing a hypothesis of symmetry in the means, viz., that the p variates have the same mean. The tests obtained by using either the extended Type I union-intersection principle or the likelihood ratio are identical and it is well known that they are equivalent to an F-test with appropriate degrees of freedom However, from the standpoint of confidence procedures, it is shown that the usual elliptical region can be replaced by simultaneous interval statements on parametric functions which are measures of departure from the null hypothesis. Also using a "truncation" procedure it is shown that one can study contrasts which are of particular interest and are components of the null hypothesis. The interval statements, which have a joint confidence coefficient $\geqq (1 - \alpha)$, are easier to use than the elliptical regions which have an exact confidence coefficient $(1 - \alpha)$. Received July 7, 1958.)

80. On Stochastic Approximation.

C DERMAN AND J. SACKS, Columbia University.

A very general th[illegible] "On stochastic approximation", Proceedings of the T[illegible]

[illegible]

vectors. The one-dimensional theorem is a consequence [illegible]

lemma. *If* $\{a_n\}$, $\{b_n\}$, $\{c_n\}$, $\{\delta_n\}$ *and* $\{\xi_n\}$ *are sequences of real numbers satisfying the following conditions*: (i) $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ *are positive*, (ii) $\{\xi_n\}$ *are non-negative*, (iii) $\lim a_n = 0$, $\sum b_n < \infty$, $\sum c_n = \infty$, $\sum \delta_n < \infty$, (iv) $\xi_{n+1} \leqq \max (a_n, (1 + b_n)\xi_n + \delta_n - c_n)$ *for all* n *greater than some* N, *then* $\lim \xi_n = 0$. The multi-dimensional theorem follows from a slightly modified version of the above lemma. (Received July 7, 1958.)

81. A Classification Problem. OSCAR WESLER, University of Michigan.

The following version of "the problem of the k-faced die" is considered: Nature's pure strategies make up two sets of states, Ω_1 consisting of the $k!$ states got by permuting a known probability distribution $p = (p_1, p_2, \cdots, p_k)$ over the faces $1, 2, \cdots, k$ of a k-faced die, Ω_2 consisting similarly of the $k!$ states arising from a known distribution $q = (q_1, q_2, \cdots, q_k)$. Classification is made on the basis of N observations given by the sufficient statistic $r = (r_1, r_2, \cdots, r_k)$ representing the number of times each face appears. Let φ be a randomized statistical decision procedure, and let $\alpha(\varphi)$ and $\beta(\varphi)$ be the maxima of the probabilities of errors of the first and second kind, respectively. Then we wish to minimize $\beta(\varphi)$ subject to $\alpha(\varphi) = \alpha_0$. The class of unique symmetric procedures φ^* optimal in this extended Neyman-Pearson sense is found by a game-theoretic, minimax method, and from the invariance of the problem under the symmetric group of permutations on k letters. A simplification is given for large N, in which the φ^* are replaced by *kaleidoscopic* tests, determined by a one-parameter family of hyperplanes and their symmetric images. Finally, it is shown that, for $k = 2$, the φ^* and the kaleidoscopic approximations are in exact agreement for every N. (Received July 7, 1958.)

82. Generalization of Palm's Loss Formula for Telephone Traffic. V. E. BENEŠ, Bell Telephone Laboratories, Inc.

Let F be a real non-negative function on a space X, let $\mathfrak{F}$ be a Borel field of X-subsets, and let ξ_k, $k = 0, 1, 2, \cdots$ be a stationary Markov process taking values in X, with transition function $p(\xi, A)$ for ξ in X and A in $\mathfrak{F}$. We interpret the numbers $F(\xi_k)$ as the inter-arrival times of telephone calls at a trunk group. There are N trunks, lost calls are cleared, and holdingtimes of trunks are independent, with a negative exponential distribution of mean, γ^{-1}. We prove the following result: If P is the stationary probability measure of ξ_k, then the chance of loss (of finding all N trunks busy) is $\left[\sum_0^N \binom{N}{n} A_n\right]^{-1} A_0 P(X)$, with $A_N = I$ and $A_n = K_N[I - K_N]^{-1} \cdots K_{n+1}[I - K_{n+1}]^{-1}$, where K_n is the operator whose action on a measure μ is defined by $K_n \mu(A) = \int X \int A \exp \{-n\gamma F(\xi)\} p(\eta, d\xi)\mu(d\eta)$. Palm's formula applies to the case $X = (0, \infty)$, $F(\xi) \equiv \xi$, ξ_k independent. Our formula has the same algebraic form as Palm's, but the multiplicative constants have been replaced by operators. The inverses indicated in our formula exist under weak hypotheses. (Received July 14, 1958.)

83. Factorial Analysis of Life-Tests. MARVIN ZELEN, National Bureau of Standards.

Consider a factorial experiment involving the factors A and B having levels a and b respectively. Let a life-test experiment be planned such that n items are tested for each of the ab factorial combinations and the test is terminated when exactly r ($r \leqq n$) of the test items have failed. Assume that the underlying distribution of failures for the (i, j) factorial combination ($i = 1, 2, \cdots, a$; $j = 1, 2, \cdots, b$) is $p(x_{ij}) = \theta_{ij}^{-1} \exp [-(x_{ij} - A_{ij})\theta_{ij}^{-1}]$ for $x_{ij} \geqq A_{ij}$, where $\theta_{ij} = m a_i b_j c_{ij}$. Maximum likelihood estimates are found for the param-

eters m, a_i, b_j, and c_{ij}. Likelihood ratio tests are given for testing various hypotheses for these parameters as well as approximations for the small sample distribution of these tests. (Received July 14, 1958.)

84. Unbiased Estimation for Functions of Location and Scale Parameters. R. F. TATE.

Integral transform theory is employed to obtain unbiased estimators (which in many cases have the minimum variance property) for functions of a location parameter θ and/or a scale parameter σ. Applications are made to the gamma distribution with parameters considered together and separately, and to truncated distributions in general. A simple formula is presented for estimating any differentiable function of a single location parameter of truncation; no calculation of distributions or conditional expectations is required in order to find a minimum variance unbiased estimator. Special attention is paid throughout the paper to the estimation of the functions $P(X \varepsilon A \mid \theta)$, $P(X \varepsilon A \mid \sigma)$ and $P(X \varepsilon A \mid \theta, \sigma)$, where A is an arbitrary Borel set. (Received July 21, 1958, revised July 25, 1958.)

85. Theory of Successive Two-Stage Sampling. (Preliminary report) B. D. TIKKIWAL (By title)

The general theory of Univariate Sampling on Successive Occasions have been studied by the author [J. Ind. Soc. Agric. Stat., Vol. 8 (1956), pp. 84–90] under a specified sampling scheme and correlation pattern. Here the sampling units selected for study on various occasions are completely enumerated. The present paper gives the best estimator and its variance under the same sampling scheme when each of the primary units (assumed of the same size M) are not completely enumerated but observed only on a sub-sample of size m. It is shown that the form of the best estimator is the same as in the univariate case when the pattern of correlation is the same at both the stages. It is further noted that, for an infinite population and $M = \infty$, the variance of the best estimator on the hth occasion is given by ϕ_h/n_h''. V in the notations of the above paper and where V is the variance of the simple two stage sampling mean when only one primary unit is selected on the hth occasion. (Received August 1, 1958.)

86. Functions of Markov Chains (Preliminary Report), MURRAY ROSENBLATT, Indiana University.

Let X_n, $n = 0, 1, \cdots$ be a Markov Chain with initial [illegible] and stationary transition probability matrix $P = (p_{ij})$ $i, j = 1, 2, \cdots$. Let [illegible] and let S_α, $\alpha = 1, 2, \cdots$, be the sets of states of X_n that [illegible] Y_n. [illegible] consist of those sets of states into which one has access with positive probability from at most one set of states. Class two is the complementary class of states. [illegible] sufficient condition that Y_n be Markovian (for a fixed [illegible]) [illegible] w_i of the X_n process, is given as follows: (i) If S_α belongs to class two, $\sum$ [illegible] C_{S_α, S_β} for all i, β. Here $p_{i, S_\alpha} = \sum_{j \in S_\alpha} p_{ij}$. (ii) Given any sequence of sets of states [illegible], $\cdots$, S_n where S_1 is of class two and $S_2, \cdots, S_{n-1}$ of class one, $\sum$ [illegible] $c^{(n-1)}_{S_{n-1}, S_n}$ for all i if there is positive probability of the path [illegible] September 8, 1958.)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of The Institute news items of interest

Personal Items

Dr. Churchill Eisenhart has been granted a Rockefeller award "in recognition of outstanding public service," and will spend the coming academic year in England, engaged in research. Dr. Eisenhart, chief of the Statistical Engineering Laboratory, which he organized in 1946, will be based at the Research Techniques Unit of the London School of Economics and Political Science, where he will continue preparation of material for a unified treatment of the fundamentals of measurement theory and practices as related to the needs of the biological, physical, social and behavioral sciences.

Ira G. Spicer, formerly Project Leader of Technical Analysis at Minneapolis-Honeywell, has taken a position as Research Engineer with the Lockheed Missile Systems Division in Sunnyvale, California.

In August, Nelson M. Blachman will take a two-year leave of absence from his job at the Sylvania Electronic Defense Laboratory, Mountain View, California, to become a Scientific Liaison Officer at the Branch Office of the U. S. Office of Naval Research in London, England, where he will carry on liaison with European scientists in the field of electronics.

Alan T. James of the Division of Mathematical Statistics, Commonwealth Scientific and Industrial Research Organization, Australia, will be a Visiting Lecturer at Yale University for the academic year 1958–1959.

Alfred Lieberman, formerly with the Bureau of Ships, has now joined the staff of the Institute for Defense Analyses, Washington, D. C.

Herman Wold has accepted an invitation to serve as Visiting Professor during the academic year 1958–1959 at Columbia University, Economics Department, New York.

Margaret P. Martin has taken the position of Associate Professor of Preventive Medicine (Biostatistics) at the Upstate Medical Center of the State University of New York at Syracuse. She formerly held a similar position at Vanderbilt University.

Hian Liang Ang, Drs. Math. completed his work for his Master's degree in Statistics at the University of California at Berkeley in October, 1957, and continued his work toward a Ph.D. degree. He goes back to Indonesia in August, 1958, to resume his post as Lecturer of Mathematics at the University of Indonesia at Bandung.

Mr. Ulysses V. Ward was appointed Instructor of Mathematics at Howard University in September, 1957.

John E. Freund has recently been appointed Chairman of the Department of Mathematics of Arizona State College at Tempe (soon to be called Arizona State University.)

Patrick Billingsley has accepted a position as Assistant Professor in the Department of Statistics of the University of Chicago.

John W. Morse has resigned from position as Head of Economics at Keuk College, New York, to teach Statistics full-time as Assistant Professor, Economics at Hobart and William Smith Colleges, Geneva, New York, doin statistical consulting and developing inventions.

John W. Mayne, Director, Operational Research, Royal Canadian Navy ha been posted by the Defense Research Board to the SHAPE Air Defense Technica Centre, The Hague, Netherlands, to be Chief of an Operational Research Section. He expects to be in Europe for about three years.

William E. Jaynes has accepted a position as an Assistant Professor of Industrial Psychology and Statistics, and Director of the Bureau of Industria Testing and Institutional Research at the University of Omaha in Omaha Nebraska.

Charles T. Lewis has recently accepted a position as operations analyst in the Operations Research Group at Convair in Fort Worth, Texas.

R. E. Beckwith has accepted a position as Senior Research Engineer with the California Institute of Technology Jet Propulsion Laboratory

Frances Campbell Amemiya has resigned her position as Chairman of the Department and Professor of Mathematics at George Pepperdine College in Los Angeles. She is now Associate Professor of Mathematics at the California Western University.

Dr. John E. Walsh, formerly with the Military Operations Research Division of Lockheed Aircraft Corporation is now with the Operations Research Group of the System Development Corporation, 2400 Colorado Avenue, Santa Monica, California.

H. W. G. Deeks has been appointed Statistician in the War Office, Whitehall, London, S.W.1.

Dr. Om P. Aggarwal has returned to Purdue University as Associate Professor after spending a year at the University of Alberta (1956–57) as Visiting Associate Professor and another at the University of Saskatchewan. While in Canada, Professor Aggarwal was also a Fellow at the Summer Research Institute of the Canadian Mathematical Congress which is held every summer at Queen's University, Kingston, Ontario

Ingram Olkin, on sabbatical leave from Michigan State University, will be at Stanford University for the academic year 1958–1959.

J. E. Morton is serving as Statistical Adviser to the UN Economic Commission for Asia and the Far East in Bangkok, Thailand; he is also giving a course at Chulaloregkorn University in Bangkok on Linear Programming.

On July 2, 1958, Alan H. Gepfert became Director of Statistical Research of the Chicago and North Western Railway Company. The major present job is to develop economic models by which to forecast revenues. Also concerned with application of sampling and regression analyses to cost-finding and general corporate planning. Mr. Gepfert was formerly a member of operations research group and faculty of Case Institute of Technology.

The Data Processing Division of International Business Machines Corpor

tion completed its move to Westchester during July. Its new address is International Business Machines Corporation, Data Processing Division, 112 East Post Road, White Plains, New York.

Professor Herbert Solomon, Teachers College, Columbia University, is spending a sabbatical year at Stanford at Berkeley. His mailing address is Statistics Dept., Stanford University, Stanford, California.

Fellowship and Research Opportunities, National Academy of Sciences—National Research Council, Division of Mathematics

The Division of Mathematics calls attention to the fact that several foundations and offices offer financial support for research in mathematics during the year 1959-60. A number of fellowships will be made available, as well as opportunities for mathematicians to engage in basic research. A partial list, with comments, is given below.

1. *National Science Foundation.* The National Science Foundation sponsors various fellowship programs in the sciences, including mathematics.

Predoctoral fellowships are awarded annually at the First Year, Intermediate, and Terminal Year levels of graduate study. Applications for 1959-1960 will be available in October 1958 from the National Academy of Sciences-National Research Council until the closing date in early January 1959; Award date—March 16, 1959.

Science Faculty fellowships for college science teachers (including mathematics) who plan to continue teaching and wish to increase their competence as teachers are at the present time offered semi-annually. Eligibility requirements include a baccalaureate degree and three (3) years of full-time experience in teaching natural science subjects at the collegiate level. Awarded annually. The program will be open from May to October. Awards will be announced in early December. Address all inquiries for information and applications to National Science Foundation, Division of Scientific Personnel and Education, Washington 25, D. C.

Postdoctoral fellowships (in making inquiry about postdoctoral awards specify program)

(1) *Regular* postdoctoral fellowships—primarily for recipients of the doctoral degree; awarded semi-annually. Program for 1959-60 concurrent with predoctoral program (see above) except that program closes in December. Information and applications will be available from the NAS-NRC. The program will also be open from July to early September 1959. Awards are announced in March and October.

(2) *Senior* postdoctoral fellowships—are open to persons who have held a doctoral degree in one of the basic fields of science for a minimum of five (5) years at time of application, or who have had equivalent training and experience. Awarded annually. Applications are available from the National Science Foundation, Division of Scientific Personnel and Education, Washington 25, D. C. The program will be open from May to October. Awards will be announced in early December.

Research Grants. The National Science Foundation also supports basic research in the mathematical sciences by means of grants. While proposals for such support are accepted at any time, individuals desiring support to begin in the summer or at the beginning of a fall semester should preferably submit their proposals in the mathematical sciences by November 1; persons desiring support to begin in the spring semester should preferably submit their proposals by May 1. Instructions for the preparation of proposals, contained in a booklet entitled Grants for Scientific Research, may be obtained upon request from the Program Director for Mathematical Sciences, National Science Foundation, Washington 25, D. C.

2. *Office of Naval Research.* The Office of Naval Research, through contracts with universities and other organizations, supports basic research in broadly selected fields of mathematics. Proposals should be directed to the Mathematics Branch, Office of Naval Research, Washington 25, D. C. In addition, postdoctoral research associateships in pure mathematics are being established under contracts with the ONR at selected universities. For details and application forms write to the above address

3. *Air Force Office of Scientific Research* The Air Force Office of Scientific Research supports research in mathematics directly through contracts with colleges, universities, foundations and industrial laboratories Such organizations are encouraged to submit proposals for research in mathematical fields in which they specialize Proposals should be mailed to the Commander, Air Force Office of Scientific Research, Attn Mathematics Division, Washington, 25, D. C.

4. *Office of Ordnance Research, U. S Army* Among the functions of the Office of Ordnance Research is the support of basic research in mathematics Proposals for projects are ordinarily made by individual scientists or groups of scientists in a form which leads to a contract between the Office of Ordnance Research and a university or research laboratory For further information write to Commanding Officer, Office of Ordnance Research, Box CM, Duke Station, Durham, North Carolina

5. *Fulbright Awards—Public Law 584 (79th Congress)* Approximately 400 awards are offered annually for university lecturing and postdoctoral research in all academic fields in Argentina, Australia, Brazil, Burma, Chile, Colombia, Ecuador, India, New Zealand, Pakistan, Paraguay, Peru, the Philippines and Thailand (competition for the preceding countries closes April 15, 1959); Austria, Belgium-Luxembourg, Republic of China, Denmark, Finland, France, Germany, Greece, Iceland, Iran, Ireland, Israel, Italy, Japan, the Netherlands, Norway, Turkey, and the United Kingdom including colonial dependencies (competition for the latter countries closes October 1, 1959). In both cases awards are for the academic year 1960-61 (the 1959-60 competition for Europe closes October 1, 1958), but in the former group of countries the academic year begins in the spring or summer instead of the autum. Awards are payable in foreign currency and usually include travel for the grantee, but not for members of his family, and a maintenance allowance, which may be adjusted in relation to the number of accompanying dependents up to four Requests for information should be addressed to the Committee on International Exchange of Persons, Conference Board of Associated Research Councils, 2101 Constitution Avenue, Washington 25, D. C

6. *National Bureau of Standards. Naval Research Laboratory Air Research and Development Command.* Postdoctoral resident research associateships are available in a variety of sciences including mathematics and are tenable at the Washington, D C. and Boulder Colorado laboratories of the National Bureau of Standards, at the Naval Research Laboratory in Washington, D. C.; and at selected development and research centers of the Air Research and Development Command. Necessary facilities and equipment incident to the research of the associate will be provided For further information write to Fellowship Office, National Academy of Sciences-National Research Council, 2101 Constitution Avenue, Washington 25, D. C. Applications for the 1959-60 program must be filed on or before January 19, 1959.

7. *Atomic Energy Commission.* The Division of Research of the Atomic Energy Commission through contracts with universities and other organizations supports research in the fields of numerical analysis, digital computer design, programming research, and re- [illegible] Division of Research, Atomic Energy

[illegible] erated by As-

[illegible] mission offers

and

Energy Commission approval is a prerequisite. The appointee may work in numerical analysis, digital computing, mathematical physics, differential equations, probability and statistics, and various specialized branches including reactor theory, hydrodynamics, and orbit theory. Computational facilities are available. Letters from candidates should give details of personal history, scientific background, and qualifications; two letters of recommendation, one from the applicant's research professor, are required. Applications should be directed to M. E. Rose, Head, Applied Mathematics Division, Brookhaven National Laboratory, Upton, Long Island, New York.

September 1, 1958

S. S. Wilks, *Chairman*
Division of Mathematics
M. H. Martin, *Executive Secretary*
Division of Mathematics

Committee on Statistics

A new committee, the Committee on Statistics, has been established in the Division of Mathematics, NAS-NRC. It has been established in the Division as the successor to the Committee on Applied Mathematical Statistics which was appointed in 1942 and placed directly under the Academy-Research Council Governing Board. The funds in the custody of the earlier committee, and amounting to approximately $5,000, have been transferred to the custody of the new committee.

Cost-Free Digital Computer Time

As announced in the March, 1958, *Annals*, pp. 343–347, the Committee on Mathematical Tables of the Institute of Mathematical Statistics has made a survey of cost-free time on digital computers in the United States. The survey, in which 171 digital computer installations were queried, is now complete. Cost-free time is available at approximately 40 installations in at least 18 states in all parts of the United States. Members of the Institute of Mathematical Statistics who wish to avail themselves of some of this cost-free time to compute on a problem of general interest, i.e., a problem which might lead to publication of results in a professional journal, are invited to get in touch with the Chairman of the Subcommittee on Cost-Free Machine Time, Professor Fred C. Leone, Statistical Laboratory, Case Institute of Technology, Cleveland 6, Ohio. Advice on the preparation of specific tables (but *not* advice on programming or numerical analysis) is available from the other subcommittees and the reader is referred to the March, 1958, *Annals* for a complete listing of them.

D. B. Owen, *Chairman*
Committee on Mathematical Tables

Preliminary Actuarial Examinations Prize Awards

The winners of the prize awards offered by the Society of Actuaries to the nine undergraduates ranking highest on the score of Part 2 of the 1958 Preliminary Actuarial Examination are as follows:

First Prize of $200: Daniel G. Quillen, Harvard University.

Additional Prizes of $100 each: Edward J. Barbeau, Jr., Toronto University, William H. Blake, Jr., George Washington University; Theodore M. Jungreis, Rensselaer Polytechnic Institute; David H. Krantz, Yale University, Joe Lipman, Toronto University; Dennis W. Moore, Harvard University; Theodore S. Rosky, State University of Iowa; Lawrence A Shepp, Brooklyn Polytechnic Institute.

The Society of Actuaries has authorized a similar set of nine prizes for the 1959 examinations on Part 2.

The Preliminary Actuarial Examinations consist of the following three examinations: Part 1. Language Aptitude Examination (Reading comprehension, meaning of words and word relationships, antonyms, and verbal reasoning.) Part 2. General Mathematics Examination. (Algebra, trigonometry, coordinate geometry, differential and integral calculus) Part 3. Special Mathematics Examination. (Probability and statistics.)

The 1959 Preliminary Actuarial Examinations will be prepared by the Educational Testing Service under the direction of a committee of actuaries and mathematicians and will be administered by the Society of Actuaries at centers throughout the United States and Canada on May 13, 1959 The closing date for applications is April 1, 1959.

Postdoctoral Study in Statistics

Awards for study in statistics by persons whose primary field is not statistics but one of the physical, biological, or social sciences to which statistics can be applied are offered by the Department of Statistics of the University of Chicago. The awards range from $3,600 to $5,000 on the basis of an eleven month residence. The closing date for application for the academic year 1959-60 is February 16, 1959. Further information may be obtained from the Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois.

Nonparametric Statistics

A revision is being made of "Bibliography of Nonparametric Statistics and Related Topics," *Journal of the American Statistical Association* 48 (1953) pp. 844–906. Material through 1959 is to be included with more emphasis, it is

Energy Commission approval is a prerequisite. The appointee may work in numerical analysis, digital computing, mathematical physics, differential equations, probability and statistics, and various specialized branches including reactor theory, hydrodynamics, and orbit theory. Computational facilities are available. Letters from candidates should give details of personal history, scientific background, and qualifications; two letters of recommendation, one from the applicant's research professor, are required. Applications should be directed to M. E. Rose, Head, Applied Mathematics Division, Brookhaven National Laboratory, Upton, Long Island, New York.

September 1, 1958

S. S. WILKS, *Chairman*
Division of Mathematics
M. H. MARTIN, *Executive Secretary*
Division of Mathematics

Committee on Statistics

A new committee, the Committee on Statistics, has been established in the Division of Mathematics, NAS-NRC. It has been established in the Division as the successor to the Committee on Applied Mathematical Statistics which was appointed in 1942 and placed directly under the Academy-Research Council Governing Board. The funds in the custody of the earlier committee, and amounting to approximately $5,000, have been transferred to the custody of the new committee.

Cost-Free Digital Computer Time

As announced in the March, 1958, *Annals*, pp. 343–347, the Committee on Mathematical Tables of the Institute of Mathematical Statistics has made a survey of cost-free time on digital computers in the United States. The survey, in which 171 digital computer installations were queried, is now complete. Cost-free time is available at approximately 40 installations in at least 18 states in all parts of the United States. Members of the Institute of Mathematical Statistics who wish to avail themselves of some of this cost-free time to compute on a problem of general interest, i.e., a problem which might lead to publication of results in a professional journal, are invited to get in touch with the Chairman of the Subcommittee on Cost-Free Machine Time, Professor Fred C. Leone, Statistical Laboratory, Case Institute of Technology, Cleveland 6, Ohio. Advice on the preparation of specific tables (but *not* advice on programming or numerical analysis) is available from the other subcommittees and the reader is referred to the March, 1958, *Annals* for a complete listing of them.

D. B. OWEN, *Chairman*
Committee on Mathematical Tables

Preliminary Actuarial Examinations Prize Awards

The winners of the prize awards offered by the Society of Actuaries to
undergraduates ranking highest on the score of Part 2 of the 1958 Prel
Actuarial Examination are as follows:

First Prize of $200: Daniel G. Quillen, Harvard University.

Additional Prizes of $100 each: Edward J. Barbeau, Jr., Toronto Uni
William H. Blake, Jr., George Washington University; Theodore M. J
Rensselaer Polytechnic Institute; David H. Krantz, Yale Universi
Lipman, Toronto University; Dennis W. Moore, Harvard University; T
S. Rosky, State University of Iowa; Lawrence A Shepp, Brooklyn Pol
Institute.

The Society of Actuaries has authorized a similar set of nine prizes
1959 examinations on Part 2.

The Preliminary Actuarial Examinations consist of the following thre
nations: Part 1. Language Aptitude Examination. (Reading compre
meaning of words and word relationships, antonyms, and verbal rea
Part 2. General Mathematics Examination. (Algebra, trigonometry, co
geometry, differential and integral calculus.) Part 3. Special Math
Examination. (Probability and statistics.)

The 1959 Preliminary Actuarial Examinations will be prepared by t
cational Testing Service under the direction of a committee of actua
mathematicians and will be administered by the Society of Actuaries at
throughout the United States and Canada on May 13, 1959. The closi
for applications is April 1, 1959.

Postdoctoral Study in Statistics

Awards for study in statistics by persons whose primary field is not st
but one of the physical, biological, or social sciences to which statistics
applied are offered by the Department of Statistics of the University of C
The awards range from $3,600 to $5,000 on the basis of an eleven mon
dence. The closing date for application for the academic year 1959–60 is
ary 16, 1959. Further information may be obtained from the Departm
Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois.

Nonparametric Statistics

A revision is being made of "Bibliography of [illegible]
Related Topics," *Journal of the American* [illegible]
pp. 844–906. Material through 1959 is to be [illegible]

hoped, on applications than previously. References (particularly to the non-English literature), reprints, and technical reports on the theory or applications of nonparametric statistics would be greatly appreciated. Also, corrections and additions to the original bibliography are desired.

University of Michigan Summer Program in Health Statistics

The School of Public Health, University of Michigan, will have a summer program in health statistics, June 18 through August 1, 1959. The faculty will be assembled from many of the schools of public health, and from the ranks of leading workers in the field of statistics in the health sciences. Tentative course titles are: Statistical Methods in Public Health, Management of Health Agency Records, Registration and Vital Statistics, Biostatistics in the Health Sciences, Demographic Methods in Public Health, Statistical Methods in Epidemiology, Sampling Techniques in the Health Sciences, Advanced Biostatistics in the Health Sciences, Statistical Methods in Biological Assay.

Further information can be obtained from F. D. Hemphill, School of Public Health, University of Michigan, Ann Arbor, Michigan.

Additional Doctoral Dissertations

The following doctorates, conferred in 1957, should be added to the list published in the June, 1958 issue of these *Annals*.

Kupperman, Morton, The George Washington University, major in mathematical statistics, "Further Applications of Information Theory to Multivariate Analysis and Statistical Inference."

McCall, Chester H., Jr., The George Washington University, major in mathematical statistics, "The Linear Hypothesis, Information, and the Analysis of Variance."

NaNagara, Prasert, Cornell, major in statistics, "Lattice Rectangle Designs."

New Members

The following persons have been elected to membership in The Institute

May 14, 1958, to June 23, 1958

Allen, (Rev.) Raymond W., Ph.D. (St. Louis University), Chairman, *Department of Mathematics, Xavier University, Cincinnati 7, Ohio.*

Amster, Sigmund J., M.S. (Columbia University), Student, University of North Carolina, *119 Harvey Street, Philadelphia, Pa.*

Burton, Ellison Stanley, B.A. (Amherst College), Systems Engineer (Statistician), Consultant, Harper Engineering Company, Santa Monica, California, *4650 East 19th Street, Tucson, Arizona.*

Cook, William H., A.B. (Hofstra College), Mathematical Statistician, *U. S. Bureau of the Census, Statistical Research Division, Washington 25, D. C.*

Edwards, Bernard, B.Sc. (London), Lecturer in Statistics, *Municipal College of Commerce, College Street, Newcastle Upon Tyne, Northumberland, England*

Friedman, Morton Philbert, M A. (Ohio State University), Student, Ohio State University, Columbus, Ohio, *885 Northwest Blvd., Columbus 12, Ohio*

Goodman, Arnold F., B.S. (N. C State College), Graduate Assistant, Stanford University, Stanford, California, *621 Harvard Avenue, Menlo Park, California*

Hakim, Muhammad A., M Sc (Univ of Calcutta) Graduate Student, *University of California, Department of Statistics, Berkeley 4, California*

Harkness, William L., M.A (Michigan State University) Special Graduate Research Assistant, *Department of Statistics, Michigan State University, East Lansing, Michigan*

Higa, Seiko, B A , (Pacific University) Statistician, *Finance Department, U. S Civil Administration of the Ryukyu Islands, Naha, Okinawa, Ryukyu Islands.*

Hill, Bruce M., M S (Stanford University), Assistant in Statistics, Graduate Student, *Stanford Statistics Department, Stanford University, Stanford California*

Johnson, Whitney Larsen, M S (University of Minnesota), Instructor, Dept of Math Institute of Technology, University of Minnesota, Minneapolis 14, Minn *2175 W Folwell, St. Paul 8, Minn*

Kaller, Cecil, M.A (Saskatchewan), Research Fellow, Purdue University, Lafayette, Indiana, *Statistical Laboratory, Purdue University, Lafayette, Indiana*

Masuyama, Motosaburo, (Doctor of Science), Chief of the Laboratory of Environmental Hygiene, Meteorological Research Institute, Tokyo, *Institute of Physical Therapy & Internal Medicine, Faculty of Medicine, Tokyo University, Bunkyo-ku, Tokyo, Japan*

Middleton, David, Ph D (Howard University), Consulting Physicist, *23 Park Lane, Concord, Mass.*

Mills, Harlan Duncan, Ph D (Iowa State) Research Associate, Princeton University, Princeton, New Jersey, *186 Elm Road, Princeton, New Jersey.*

Nemenyi, Peter B., M A. (Princeton University), Assistant Statistical Analyst, *Metropolitan Life Insurance Company, Madison Avenue, New York 10, N Y* , also Lecturer, Hunter College, Pk Ave, and 68th Street, New York

Niedzielski, Edmunk L., Ph.D. (Fordham University), Research Chemist, *E I DuPont De Nemours and Co , Petroleum Laboratory, P O Box 1671, Wilmington 99, Del*

Okamoto, Masashi, M S. (Tokyo University), Lecturer in Mathematical Statistics, Osaka University, Japan, *Nakanoshim, Kita-ku, Osaka, Japan*

Pincus, Louis, B S , (The City College of New York), Senior Statistician, New York City Department of Health, 125 Worth Street, New York 12 N Y , *451 Kingston Avenue B-7, Brooklyn 25, New York*

Randels, Robert B., Ph D (University of Michigan) Physicist, *Corning Glass Works, Houghton Park, Corning, New York*

Robinson, Enders A., Ph D., (Massachusetts Institute of Technology), Assistant Professor of Statistics, Michigan State University, *Department of Statistics Michigan State University, East Lansing, Michigan*

Sato, Sokuro, (Tokyo Technical College) Assistant Professor, Faculty of Education, Saga University, Saga City, Saga Prefecture, Japan. *Dokkoo-koji, Mizugaemachi, Saga City, Saga City, Japan.*

Slud, Maurice H., M.A (Columbia University), Mathematical Analyst, *General Electric Company, Missile and Ordnance Systems Department, 3193 Chestnut Street, Philadelphia 4, Pennsylvania.*

Smith, William Roger, M.S (University of Wisconsin), Student, University of California, *5345 Zara Avenue, Richmond, California.*

Tsutakawa, Robert K., M S (University of Chicago), Quality Analyst A, Quality C.

Department, Pilotless Aircraft Div., Boeing Airplane Co., Seattle, Washington, *1103½ 3rd Avenue, Seattle, Washington.*

Welch, Peter D., M.S. (University of Wisconsin), Staff Engineer, *IBM Research Center, Yorktown Heights, New York.*

Zacks, Shelemyahu, B.A. (Hebrew University), *Statistician of the Building Research Station, Technion, The Technion Research Institute, Haifa, Israel.*

Zehna, Peter W., M.A. (University of Kansas) Research Assistant, Stanford University, *Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.*

REPORT OF THE CAMBRIDGE, MASSACHUSETTS MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The seventy-eighth meeting of The Institute of Mathematical Statistics and the twenty-first annual meeting was held at the Massachusetts Institute of Technology, Cambridge, Massachusetts, on August 25–28, 1958, in conjunction with the meetings of the American Mathematical Society, the Mathematical Association of America, the Society for Industrial and Applied Mathematics, and the Econometric Society.

The program of the meeting was as follows:

MONDAY, AUGUST 25, 1958

9:00 A.M. Invited Papers on Regression and Analysis of Variance

Chairman: Franklin A. Graybill, Oklahoma State University

1. *Variance Component Analysis in Models Where Effects Are Time Variables*, A. W. Wortham, (presented by Leroy Folks) Texas Instruments Inc., Dallas
2. *Industrial Experience with 2^{p-q} Fractional Factorial Experiments*, Cuthbert Daniel, New York City
3. *Confidence and Significance Procedures for Non-linear Models*, M. B. Wilk, Bell Telephone Laboratories, Murray Hill

11:15 A.M. Wald Lecture I

Chairman: J. L. Hodges, Jr., University of California, Berkeley

The Mathematical Basis of Fiducial Inference, John W. Tukey, Princeton University

2:00 P.M. Invited Papers on Estimation and Testing

Chairman: Donald L. Burkholder, University of Illinois

1. *Power of the Chi-square Test*, J. L. Hodges, Jr., University of California, Berkeley
2. *On Solutions of Dorfman's Mass-Testing Problem*, Milton Sobel, Bell Telephone Laboratories, Allentown
3. *Linear Regression in the Multivariate Normal Case*, Charles Stein, University of California, Berkeley

4:00 P.M. Invited Papers on Testing

Chairman: D. B. Owen, Sandia Corporation

1. *Partial Orderings of Probabilities of Rank Orders*, I. Richard Savage, University of Minnesota
2. *Simple Methods for Analysis of Two-Action Problems with Linear Costs*, Robert Schlaifer, Harvard University

Chairman: J. R. Blum, University of Indiana

1. *A Moment-Problem with Restriction on Smoothness*, C. L. Mallows, Princeton University
2. *About the Central Limit Problem*, Michel Loève, University of California, Berkeley
3. *Hausdorff Dimension and Information Theory*, Patrick Billingsley, Princeton University and University of Chicago
4. *A Geometry of Binary Sequences Associated with a Class of Error-Correcting Codes*, Roy R. Kuebler, Jr., University of North Carolina

4:00 P.M. Special Invited Address

Chairman: Shanti S. Gupta, University of Alberta

Multiple Decision Selection Procedures, Milton Sobel, Bell Telephone Laboratories, Allentown

WEDNESDAY, AUGUST 27, 1958

9:00 A.M. Invited Papers on Sequential Analysis

Chairman: Kenneth J. Arnold, Michigan State University

1. *A Modification of Sequential Analysis to Reduce the Sample Size*, T. W. Anderson, Center for Advanced Study in Behavioral Sciences and Columbia University
2. *Binomial Sequential Testing*, Colin R. Blyth, Stanford University
3. *Unbiased Sequential Estimation for Binomial Populations*, Morris H. DeGroot, Carnegie Institute of Technology

11:15 A.M. Wald Lecture III

Chairman: Marvin Zelen, National Bureau of Standards

The Interpretation of Fiducial Inference, John W. Tukey, Princeton University

2:00 P.M. Invited Papers on Probability and Stochastic Processes II. (Simultaneous with Invited Papers on Random Balance.)

Chairman: Max Woodbury, New York University

1. *Semigroups of Operators and Stochastic Processes*, A. V. Balakrishnan, University of California, Los Angeles
2. *Independent Polynomials in Normal Variates*, R. G. Laha, Catholic University of America and Columbia University
3. *On Multi-event Renewal Processes*, Ronald Pyke, Stanford University

2:00 P.M. Invited Papers on Random Balance

Chairman: Frank J. Anscombe, Princeton University

1. *Introductory Remarks*, Frank J. Anscombe, Princeton University
2. *On the Analysis of Screening Experiments*, E. M. L. Beale, Princeton University, and C. L. Mallows, Princeton University
3. *Analysis Methods for Randomly Balanced Factorial Designs*, A. P. Dempster, Bell Telephone Laboratories and Harvard University
4. *Mathematical Outline of Polyvariable Analysis* (including *Random Balance*), F. E. Satterthwaite, Statistical Engineering Institute, Wellesley Hills

4:00 P.M. Wald Lecture IV

Chairman: M. B. Wilk, Bell Telephone Laboratories

What Importance Should We Place on Fiducial Inference? John W. Tukey, Princeton University

5:30 P.M. Business Meeting

8:00 P.M. 1959 Council Meeting

THURSDAY, AUGUST 28, 1958

9:00 A.M. Contributed Papers III (Simultaneous with Contributed Papers IV)

Chairman: ROBERT A. WIJSMAN, University of Illinois

1. *On a Limiting Distribution Due to Renyi*, D G CHAPMAN, University of Washington
2. *Stochastic Models for the Electron Multiplier Tube*, EDWARD K. DALTON, WILLARD D. JAMES, AND HOWARD G TUCKER, University of California at Riverside
3. *On Stochastic Approximation*, C. DERMAN AND J. SACKS, Columbia University
4. *Single Server Queuing Processes with a Finite Number of Sources*, GERALD HARRISON, The Teleregister Corporation
5. *A Model for Failure Data and its Applications*, ANDRE G. LAURENT, Wayne State University
6. *Generalization of Palm's Loss Formula for Telephone Traffic*, V E BENES, Bell Telephone Laboratories, Murray Hill
7. *The Moments of the Maximum of Partial Sums of Independent Random Variables*, JOHN S WHITE, Minneapolis Honeywell Regulator Co
8. *Stochastic Models for Length of Life*, BENJAMIN EPSTEIN, Wayne State University and Stanford University (By title)
9. *Tests for the Validity of an Exponential Distribution of Life*, BENJAMIN EPSTEIN, Wayne State and Stanford Universities, (by title)

9:00 A.M. Contributed Papers IV

Chairman: M. V. JOHNS, JR , Stanford University

1. *Estimation of the Medians for Dependent Variables*, OLIVE JEAN DUNN, University of California, Los Angeles
2. *The Use of Sample Quasi-Ranges in Estimating Population Standard Deviation*, H. LEON HARTER, Wright Air Development Center
3. *Order Statistics and Estimation*, M. M. RAO, University of Minnesota (by title)
4. *A Note on Order Statistics and Stochastic Independence*, GERALD S ROGERS, University of Arizona
5. *Aids for Fitting the Pearson Type III Curve by Maximum Likelihood*, J ARTHUR GREENWOOD, Iowa State College AND DAVID DURAND, Massachusetts Institute of Technology
6. *On the Relationship Algebra and the Association Algebra of the Partially Balanced Incomplete Block Design*, JUNJIRO OGAWA, University of North Carolina (by title)
7. *Factorial Analysis of Life-Tests*, MARVIN ZELEN, National Bureau of Standards
8. *On the Analysis of Factorial Experiments without Replication*, ALLAN BIRNBAUM, Columbia University (by title)
9. *On Logistic Order Statistics*, ALLAN BIRNBAUM, Columbia University (by title)
10. *Statistical Theory of Some Quantal Response Models*, ALLAN BIRNBAUM, Columbia University (by title)
11. *Optimum Designs in Regression Problems*, J. KIEFER AND J. WOLFOWITZ, Cornell University, (by title)
12. *On the Bounds for the Variance of Mann-Whitney Statistic*, JAGDISH SHARAN RUSTAGI, Michigan State University, (by title)
13. *A Characterization of Triangular Association Scheme*, S. S. SHRIKHANDE, University of North Carolina (by title)
14. *A Problem in Two-Stage Experimentation*, DONALD L. RICHTER, University of North Carolina (by title)

11:15 A.M. Special Invited Address

Chairman: SAMUEL W. GREENHOUSE, National Institutes of Health

Estimation Methods in Multivariate Analysis, EVAN J. WILLIA
State College

2:00 P.M. Special Invited Address

Chairman: MORRIS H. HANSEN, Bureau of the Census

On a Formal Structure of Professional Practice in Sampling, W. EDWARDS DEMING, New York University

3:15 P.M. Invited Papers on Mixed Topics

Chairman: H. A. DAVID, Virginia Polytechnic Institute

1. *The Number of Occupied Cells of a Particular Subclass (When Objects are Assigned to Cells at Random)*, HOWARD L. JONES, Illinois Bell Telephone Company, Chicago
2. *Statistical Theory of Tests of a Mental Ability*, ALLAN BIRNBAUM, Columbia University
3. *Properties of Some Control Chart Tests for Detecting Shifts in a Process Average*, S. W. ROBERTS, Bell Telephone Laboratories, New York.

REPORT OF THE PRESIDENT FOR 1958

To the membership of the IMS
Dear Friends:

I am writing to you shortly before my departure from the country, because I shall unfortunately not be here to address you personally at the annual meeting.

The presidency of the Institute is an honor greater than I had expected to receive in my whole life. I thank you all for your expression of recognition and confidence. I also thank for myself, and for all of us, the many officers, committee members, and representatives who have so loyally and competently handled the Institute's increasingly complex and serious business.

The president's job is not among the harder ones in the Institute, but I have found it interesting and instructive. The office entails some decisions and offers opportunities to make suggestions throughout the Institute from a central vantage point. It also offers opportunities to make mistakes, and I have made some. Those that have thus far come to light are mostly small and rectifiable.

Since the meeting at Atlantic City a year ago there have been regional meetings at Los Angeles, Gatlinburg, and Ames.

T. E. Harris announced a year ago that he wanted to relinquish the Editorship at the expiration of his term on July 1, 1958. An ad hoc committee consisting of William Cochran (Chairman), T. W. Anderson, M. S. Bartlett, T. E. Harris, W. A. Wallis, and S. S. Wilks recommended to the Council that W. H. Kruskal be appointed to the Editorship. The Council has accepted this recommendation and Kruskal has accepted the appointment. The *Annals* continues to grow and improve by leaps and bounds, as I expect you will hear in detail from the retiring Editor.

The Council decided in Atlantic City that, to meet the rising costs of printing a larger *Annals* at higher rates for printing, we ought to apply to the National Science Foundation for a grant of money for a three-year period. I hope

that it will be possible to announce along with the reading of this letter that the grant has been made.

The Council has before it a plan, submitted by an ad hoc committee, to approach the National Science Foundation for aid in preparing translations of Russian statistical and probabilistic literature.

A summer institute on nonparametric methods will be held at Minneapolis before this letter is read, and an advisory committee will report to the Council at this meeting on what plans, if any, should be made for a summer institute in 1959.

The final duty of the president is to appoint a new nominating committee and I herewith appoint: T. E. Harris, Chairman, Herman Chernoff, David Cox, J. C. Kiefer, and W. J. Youden.

Deeply regretting having to take my leave this way, I am

Most sincerely yours,
[illegible]
President

IMS OFFICERS, COMMITTEES, AND REPRESENTATIVES

combe, R. J. Bose, D. L. Burkholder, D. G. Chapman, Cuthbert Daniel, T. S. Ferguson, Evelyn Fix, F. A. Graybill, H. O. Hartley, P. J. McCarthy, Howard Raiffa.

IMS PROGRAM COMMITTEE OF THE CENTRAL REGION: Jack Silber, Virgil Anderson, Charles Bell, H. T. David, F. A. Graybill, E. R. Immel, Bernard Ostle, M. B. Wilk (ex officio), W. Kruskal (ex officio), R. N. Bradt, Frank Graybill, Irving Burr, John Gurland, Howard Jones, Fred Leone, Boyard Rankin, Paul Rider, Jack Silbert (ex officio), Martin Wilk (ex officio).

IMS EASTERN REGION PROGRAM COMMITTEE: Boyd Harshbarger, Ralph Bradley, B. G. Greenberg, D. G. Horvitz, Carl Kossack, Herbert A. Meyer, John Pratt, Dorothy Gilford (ex officio), Martin Wilk (ex officio), W. H. Horton, J. D. Hromi.

IMS WESTERN REGION COMMITTEE: Richard Link, Fred Andrews, Charles Bell, Tom Ferguson, John Gilbert, John Hofmann, Marion Sandomire, David Stoller, Robert Tate, J. R. Vatnsdal, Frank Massey.

IMS COMMITTEE FOR SPECIAL INVITED PAPERS: W. S. Connor, G. E. P. Box, Kai-Lai Chung, W. Kruskal, G. E. Noether, M. B. Wilk, T. E. Harris member ex officio (as Editor).

IMS AD HOC COMMITTEE ON HIGH SPEED COMPUTING: A. S. Householder, G. S. Acton, R. L. Anderson, K. J. Arnold, C. F. Kossack, W. H. Kruskal, W. J. Merrill, H. A. Meyer, J. Moshman, H. W. Norton, G. J. Resnikoff, R. Slimak, Z. Szatrowski, D. Teichroew.

IMS SUBSCRIPTIONS COMMITTEE: Edward Coleman, Joe Adams, K. A. Bush, Lila Elveback, Harry Harman.

IMS COMMITTEE ON MATHEMATICAL TABLES: D. B. Owen, G. P. Steck, Paul Cox, Fred C. Leone, John W. Tukey, Marvin Zelen, R. L. Anderson, A. H. Bowker, E. E. Cureton, W. J. Dixon, C. W. Dunnett, Churchill Eisenhart, J. A. Greenwood, H. O. Hartley, William Kruskal, Daniel Teichroew, M. A. Woodbury, L. J. Savage (ex officio), J. Wolfowitz (ex officio).

IMS AD HOC COMMITTEE ON ANNUAL MEETING POLICY: Leo Katz, Cecil Craig, Churchill Eisenhart, Robert Hooke, Henry Scheffe, Martin Wilk.

AD HOC COMMITTEE TO INVESTIGATE THE POSSIBILITY OF BILLING FOR PUBLICATION IN THE *ANNALS*: A. M. Mood, A. H. Bowker, John Curtiss, Dorothy Gilford, W. Kruskal.

IMS BLACKBOARD COMMITTEE: Irving W. Burr, Herbert Robbins, Martin Wilk.

IMS EDITOR HUNT COMMITTEE: W. G. Cochran, T. W. Anderson, M. S. Bartlett, T. E. Harris, W. A. Wallis, S. S. Wilks.

IMS COMMITTEE ON RUSSIAN TRANSLATIONS: Ingram Olkin, Eugene Lukacs.

COMMITTEE TO ADVISE ON A POSSIBLE SUMMER INSTITUTE FOR 1959: David Blackwell, H. O. Hartley, Richard Savage, David Wallace, Max Woodbury.

IMS REPRESENTATIVE TO AAAS: Harold Hotelling.

AMERICAN STANDARDS ASSOCIATION COMMITTEE ON STATISTICAL NOMENCLATURE IMS REPRESENTATIVE FOR 1957–58: Max Halperin.

IMS REPRESENTATIVE IN DIVISION OF MATHEMATICS NATIONAL RESEARCH COUNCIL: W. Allen Wallis.

REPRESENTATIVES TO CONFERENCE ORGANIZATION OF THE MATHEMATICAL SCIENCES: Joseph F. Daly for 1957–58, H. B. Mann for 1957–59.

REPORT OF THE SECRETARY FOR 1958

During the past year The Institute has held its 75th through 78th meetings. A business meeting was held during the 78th (21st Annual) meeting. The Program Committees are to be congratulated on the excellent programs which

I take this opportunity to thank the many people whose hard wo so valuable to the *Annals* during my editorial term, now ending. A li who refereed papers during 1958 will be printed in an early issue. I ticularly Ann Greene, Jeanette Hiebert, Dorothy Stewart, Helen and Margaret Wray, who have carried on the work of the editoria I want to express my appreciation to The RAND Corporation for sible my work on the *Annals*.

T

E

PUBLICATIONS RECEIVED

Introduccion a La Investigacion Operativa, Instituto de Investigaciones Est rrano 123, Madrid, Spain.

Anuario Estadistico de Espana, Presidencia del Gobierno, Instituto Nacional de Ferraz 41, Madrid, Spain.

Integrals of Airy Functions, National Bureau of Standards Applied Mathemat issued May 15, 1958, 28 pp., 25 cents. (Order from the Superintendent of U.S. Government Printing Office, Washington 25, D. C.)

Table of Natural Logarithms for Arguments Between Five and Ten to Sixteen De National Bureau of Standards Applied Mathematics Series 53, issued Ma $4.00. Supersedes Mathematical Table 12. (Order from Superintendent of U. S. Government Printing Office, Washington 25, D. C.)

Feinstein, Amiel, *Foundations of Information Theory*, McGraw-Hill Book C West 42nd Street, New York 36, New York. 137 pages, $6.50.